

| | | | |
|----|-------|---|-------|
| Qy | 2011 | ProProAlaProProAlaSerAlaSerAsp-----ProHis | 2022 |
| Db | 9178 | CCCCAGGGACACCTTGATCTCTCAGAGCCGAGCACTACGCCACCGTGAGCGGCCAC | 9237 |
| Qy | 2023 | ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly | 2042 |
| Db | 9238 | CGGATCCAGCGGCACGGCTCTCTCACCACGAGCAACTGCTGGCA----- | 9282 |
| Qy | 2043 | TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerPro | 2062 |
| Db | 9283 | -----CCCCACATGTGAGCACACCGCCACGACACCCACAGTCCACAGCTCCAAAGCCA | 9336 |
| Qy | 2063 | SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeu | 2082 |
| Db | 9337 | CTCCCTCTCTCAGTCCAGGGACTG-----CAACTGCGCTTCCAGCACTGA | 9381 |
| Qy | 2083 | GluGlyGlu-----LeuArgProLysGlnProGlyProVal | 2094 |
| Db | 9382 | GAAGCAGCGCCACACACACAGTACAGCTTTTACAGCCATCCCTCTCTCCCTGG | 9441 |
| Qy | 2095 | LysLeuGlyGlyGluAlaHisLeu-----ProHisLeuArgProLeuProGluSerGln | 2113 |
| Db | 9442 | GCACCCTGGACCCCGCTATCAGACACCACCCACGGGCCA-----CCATGTCCACAG | 9498 |
| Qy | 2114 | ProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValVal | 2133 |
| Db | 9499 | CCACACCTCTCTCACTCCA-GAGACTGTC----- | 9527 |
| Qy | 2134 | ThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr-ArgHisHisProGln | 2153 |
| Db | 9528 | -----CACACCTTCCAGTGGCTTACCACCCACCGCCACACACCGGGGCCACC | 9575 |
| Qy | 2153 | nGlnLeuSerAlaProLeuProAlaPro-----LeuTyrSerPheProGlyValaserCy | 2171 |
| Db | 9576 | GGCTCTGTGGCACCCCTCTCCACCCAGGACAGCTCACATACCDAAGTGCCTACT | 9635 |
| Qy | 2171 | sProValLeuAspLeuArgProProSerAspLeuTyrLeuProProAspHisGln | 2191 |
| Db | 9636 | ACCACAACACGGGGTTCCAGCGCCC-----CCTCTCCAGCCCGG | 9680 |
| Qy | 2191 | Y----- | 2191 |
| Db | 9681 | ACGGCACTACGGCTCAGTGTGATGACACACACCCACACCCACCAACACCAACACC | 9740 |
| Qy | 2192 | -----AlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGln | 2208 |
| Db | 9741 | ACAACCACTGGCTCCAGGTGACCCCTCTCCATCCAGGAGCACCCACACCCGCCAGA | 9800 |
| Qy | 2208 | u-----ProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSe | 2226 |
| Db | 9801 | GTGCTGACCACC-ACCACCACCACTGTGGCCACTGGTTCTATGGCAACACCCCTCTCTAG | 9859 |
| Qy | 2226 | rProProGluGlyMetThrCluProGlyHisSerArgSerAlaValTyrProLeuLeuTy | 2246 |
| Db | 9860 | CACACAGACAGTGGTACTCCCCATCATGACCACCGGCCACT-----ACATCAC | 9913 |
| Qy | 2246 | rArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSe | 2266 |
| Db | 9914 | GGCCACGGGTCCACCAACCAACCC-----TCTCAACTCCAGGCACACACC | 9961 |
| Qy | 2266 | rgInProPro-----AlaPhePheSerLysLeuTh | 2276 |
| Db | 9962 | CATCCCCCAGTGTGACCAGCATGGCCACCACCCCGCAGCCACCGAGTCCCAAGGCCAC | 10021 |
| Qy | 2276 | rgLusSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLys----- | 2292 |
| Db | 10022 | TTCTCTCTCAGTCCAAGGAGTGC-AACCAACCTTCCAGTGTGTGAACAAGCACAGCCACA | 10080 |
| Qy | 2293 | -----LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGln | 2308 |
| Db | 10081 | AATCCACAGTACAGCTTTACACCCATCCCTCTCTCCACCTGTGGAGCAGCTGGAGCCG | 10140 |
| Qy | 2308 | nProGlyThrGluIlePheAsn-MetProAlaIleThrGlyThrGlyLeuMetThrTyrA | 2328 |

Search completed: April 16, 2004, 02:38:23
Job time : 14964 secs

| | | | |
|----|------|--|------|
| Db | 7344 | GTCCACACCTCCACAGTGCTTACACACCGCCACACACCGGGGCCACCGGCTCTGTG | 7403 |
| Qy | 1384 | -ProProProProProSerArgAspLeuThr | 1394 |
| Db | 7404 | GCACACCCCTCTCCACCCAGGACACAGCTCACACTACCAAAAGTGGCGACTACCAACACC | 7463 |
| Qy | 1395 | -----GluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHi | 1411 |
| Db | 7464 | ACGGGCTTCACAGTCACC-CCCTCTCCACCGCCAGGACCGCACCGCTCCAGTGTG | 7522 |
| Qy | 1411 | sGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluLeuProArgG1 | 1431 |
| Db | 7523 | GATCAGCACACACACACACACACAGTGGCTCCACGGTGACCCCTCTCCATCC | 7582 |
| Qy | 1431 | uGluLeuArgHisThrProGlu-----LeuProLeuAlaProArgProLeuLysGluGlySe | 1450 |
| Db | 7583 | GGGGACACACACACACCCACAGTGTGACACACACACCAACCTGTGGCCACTGGTTC | 7642 |
| Qy | 1450 | rIle-----ThrGlnGlyThrProLeuLysTyrAspThrG1 | 1462 |
| Db | 7643 | TATGGCAACACCTCTCTAGCACACACAGAGTGGTACTCCCTCCATCATGTATCACAC | 7702 |
| Qy | 1462 | YAlaSerThrThr--GlySerLysLysHisAspValArgSerLeuLysLeuLysGlySerProGly | 1481 |
| Db | 7703 | GGCCACTACGATCAGCGCCACCGCTCCACAC---CAACCCCTCTCAACTCCAGGAC | 7759 |
| Qy | 1482 | ArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArg | 1501 |
| Db | 7760 | AACACCTATCCCCC-----AGAGCTGACACACACACCGC---CACAC | 7798 |
| Qy | 1502 | AlaCys-----TyrGluGluSerLeu-----LysSerArgProGlyThr | 1514 |
| Db | 7799 | ACCTGACACACACAGCAGCAGTACCTCTCTGCTGCTAGGACACACACACACAC | 7858 |
| Qy | 1515 | AlaSerSerSer--GlyGlySerIleAlaArgGlyAlaProValIleValProGluLeu | 1533 |
| Db | 7859 | CCAGTGGCGAACACACCGCCACACACACACCGGGGATCCCTGTC----- | 7903 |
| Qy | 1534 | GlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla | 1548 |
| Db | 7904 | -----CCCGCAGCAGTCCCA-----CACGGTGGCGACAGCTCGACTTC | 7942 |
| Qy | 1549 | -----ProPheAlaGlyHisLeuProArgGlySerPro | 1559 |
| Db | 7943 | GGCCACCTCAGCAGCTTGGGACACACCAACA-TCACAGAGCTTCCACGGGACTTCCC | 8001 |
| Qy | 1560 | ValThrMetArgGluPro-----ThrProArgLeuGlnGlySer-LeuSe | 1575 |
| Db | 8002 | ACACCCACGACCAACACCGGTACCAACACAGCAGCTCGACTCCAG-----CCCTGTCCA | 8055 |
| Qy | 1575 | rSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSe | 1595 |
| Db | 8056 | GCCTCTACCTTAGCAGCAGGACCCAGTACCCCTTCCCGAGGACGACCCCGG | 8115 |
| Qy | 1595 | rProHisSerThrValProGluHisHisProHis----- | 1606 |
| Db | 8116 | GCACACACGCGCCACTCCAGCAGCACCGCCACCGCCACACCCAGAACGCCGCACT | 8175 |
| Qy | 1607 | -----ProIleSerProTyrGluHis----- | 1613 |
| Db | 8176 | CGACCTGTGTCGCCACGACCAATCGGCCCCCAATACCAACGAGTGTGACCGGCT | 8235 |
| Qy | 1614 | -----LeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisI1 | 1628 |
| Db | 8236 | GTGAGCCCGAGTGTGCTGTGAGA---GTGGCTGGACTACAGCTACCCCATCGCGGGC | 8292 |
| Qy | 1628 | eProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaA1 | 1648 |
| Db | 8293 | -CCTCTGGCGGGACTTTGACACCTACTCCAAATCCGTG----- | 8331 |
| Qy | 1648 | aTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTy | 1668 |
| Db | 8332 | -----CGGCGGAGGGCGCTGTGTGAGCAGCGCCCTG | 8363 |

Qy 844 LysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLys 863
Db 5416 GGACCACTGGATCTCCACAGAGCCAGCACT-ACAGCCACCGTGACGGTCCACCGGA 5474
Qy 864 SerGluCysThrGluGluAlaGluGluClyProAlaLysGlyLysAspAlaGluAla 883
Db 5475 TCCACGGCCACC-----GCTCCTCCACC 5498
Qy 884 GluAlaThrAlaGluGlyAlaLeuLys-----AlaGluLysLysGluGly 898
Db 5499 CGGGCACTGCT---GGCACCTCAAGTGCTGACAGCAGCCAGCCACCCACAGTC 5555
Qy 899 GlySerGlyArgAlaThrAlaLysSerSerGlyAla-----ProGlnAsp 914
Db 5556 ATCAGTCCAGAGCACTCCTCCTCCAGTCCAGGACTGCAACCGCCCTTCCAGCACTG 5615
Qy 915 SerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyAspLys 934
Db 5616 AGAAGCACAGCCACACACCCACAGCTACCAGCGTTACGCCATC----- 5660
Qy 935 AsnArgLeuLeuSerProArgProSerLeu-----LeuThrProThr 948
Db 5661 -----CCTCTTCTCCTCGGCACCGCCTGGACCCCGCTATCACAGCC 5705
Qy 949 GlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp-LeuLysGlnLeuLysG1 968
Db 5706 ACCACACCCAGCCACCAGTGTCCACAGCCACACCCCTCTCCACTCCAGAGCTGTCCAC 5765
Qy 968 nArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAs 988
Db 5766 ACCTCCACAGTGTCTTACCACA----- 5787
Qy 988 pAlaAlaProThrLysProAlaPro-----ProAlaProProProGlnAsn-- 1004
Db 5788 -----CGGCCACCAACACAGGACCGGCTCTGTGGCCACCCCTCTCCACCCAGGAC 5843
Qy 1005 -----LeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSe 1023
Db 5844 GCTCACACTACCAAGTGCAGACTACCAACACAGCGGCTTCCAGCCA----- 5892
Qy 1023 rArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuPr 1043
Db 5893 -----CCCCCTCTCCAGCCACAGG-ACGGCACTCAGC----- 5924
Qy 1043 oGlyAspProProCysTriThrSerGlyLeu-ProPheProValPro-----ProArgG 1061
Db 5925 -----CCTCCAGTGTGGATCAGCAACACCAACACCAACCAACAGAGGCTCCAGGT 5977
Qy 1061 luValLysAlaSerProHisAlaProAspProSer----- 1073
Db 5978 GACCCCTCTCCATCCCGGGACCAACCCACCGCCACAGTGTGACCAACCAACCCAC 6037
Qy 1074 -----AlaPheSerTyrAlaProProGlyHis-----ProLeuProLeuG 1087
Db 6038 AACTGTGGCCACTGGTTCATGGCAACACCCCTCCTTAGCACACAGACAGTGGTACTCC 6097
Qy 1087 lyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPro- 1106
Db 6098 CCATCACTACCAACCAACCGCCACTAGCATCAGCCACCG-----GCTCGACCAACAA 6151
Qy 1107 ProLeuLysSerAla-----LysHisProSerValLeuGlu-----ArgGlnIle 1122
Db 6152 CCCTCTCAACTCCAGGGACAACCTCCCATCCCGCAGTGTGACCAACCAACCGCCAC 6211
Qy 1123 GlyAlaIleSerGlnGlyMetSerValGlnLeu-----HisVal 1135
Db 6212 ACCTGCAGCCACCAAGCAGCAGTACTCCTCTCTGCTAGGGACCAACCCACAC 6271
Qy 1136 ProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuPro 1155
Db 6272 CCAGTCCCAACACCA-----GGCCACCAACACCGGGCGTCTCTGCCCC 6319
Qy 1156 MetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLeuSerProArg 1175

Db 6320 CAGCAGTCCCCACACCG-TGCCACACAGCTGGACTTCGG-----CCACCTCGG 6366
Qy 1176 GlyGlnAlaGlyProProGluSer-----LeuGlyValProThrAlaGlnGlu 1191
Db 6367 GCATCTTGGGCACACCCACATCAGAGCCTTCCACGGGACTTCCACACCCACCA- 6425
Qy 1192 AlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGly 1211
Db 6426 GCAACACC-----GGTACCACCCAGCCCTCGACTCCA-----GCCCTTCCAGCCCT 6473
Qy 1212 IleProSerThrArgVal-----ProSerAspSerAlaIleThr----- 1224
Db 6474 CACCTAGCAGCAGGACCAACCGAGTCAACCCCTTCTCCAGGACGACCAACCCCGGCCAC 6533
Qy 1225 TyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThr 1244
Db 6534 ACCAGGGGC-----ACCTCCAGGACCAACAGCCACACCCAGCAAGACCCGACC 6587
Qy 1245 ArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuPro 1264
Db 6588 TCACACCTGTGCCAGCAGCCCA-----TGGGCCCC 6623
Qy 1265 LysGlyHisValIleTyrGluGlyLysGlyHis-ValLeuSerTyrGluGlyMe 1284
Db 6624 ATAACACAGGTGTGTACACCGGCTGTGAGCCCGAGTGTGCTGTCAGAGTGGCTGGAC 6683
Qy 1284 tSerValThrGlnCysSer----- 1290
Db 6684 TACAGTACCTCCATGCCGGGGCCCTCTGGCGGAGCTTTCGACACCTACTCAACATCCGT 6743
Qy 1291 -LysGluAspGlyArgSerSerGlyProHisGluThrAlaAlaProLysArg-- 1309
Db 6744 GCGGCCGAGGGGAGTCTGTGAGCAGCCCTCGGCTTGTGAGTGGCTGCCAGGCCAG 6803
Qy 1310 -----ThrTyrAspMe 1313
Db 6804 CTGTGTCTCCCTCGGGAGTTGGGCGAGTGTGGAATGATGAGCTTGGGCTG 6863
Qy 1313 tMetGluGlyArgValGlyArg-----AlaIle 1322
Db 6864 GTCTGAGGAAACCGTGAGCAGGTGGGAGTTCAGATGTGTTCAACTATGAATCCGT 6923
Qy 1322 eSerSerAlaSerIleGluGly----- 1329
Db 6924 GTGTTCTGTGCACTACGGCCACTGCCACGACCCCGGCCACAGCTCTACGGCCACG 6983
Qy 1330 -----LeuMetGlyArg----- 1333
Db 6984 CCCTCCTCAACTCCGGGACGACTGGATCCTCAAAAGCTGACCAACAGCCACTACG 7043
Qy 1334 -----AlaIleProGluArgHisSerProHisIleLysGluG1 1348
Db 7044 ACTGAGTCCACTGGATCCAGGCCACCCCGTCTCCACCCAGGACCACTGTGGCACC 7103
Qy 1348 nHis-----HisIleArgGly-----Se 1354
Db 7104 CCATGTGAGCACCAGCCACGACCAACCCACAGCTACAGCTCCAAAGCCACTCCCTTC 7163
Qy 1354 rIleThrGlnGlyLe-----ProArgSerTyrValGluAlaGlnGluAspTyrLeuAr 1372
Db 7164 TCCAGTCCAGGAGTGCACCGCCCTTCCAGCACTGAGAAGCAGCAGCCACCAACCA 7223
Qy 1372 gArgGluAla----- 1375
Db 7224 GTTACAGCTTTTACAGCATCCCTCTCTCTCCCTGGGACCACTGGAGCCCGCTATCA 7283
Qy 1376 -----LysLeuLeuLysArgG1 1381
Db 7284 CAGACCAACCAACCCAGCCACCATGTTCACAGCCACACCTCTCCACTCCAGAGACT 7343
Qy 1381 uGlyThr----- 1383

QY 1798 gAspArgAspArgGluArgAspArgAspArgGluArgGluArgSerIleLeuThrSerTh 1818
Db 77106 -----ACGTGCTCAGCGACGCGCGTTCGACGATGATGTTTCGGCGAA 77060
QY 1818 rThr---ThrValGluHis-----AlaPro----- 1825
Db 77059 CACGGTGGCGTACGACATAACTGGGCTCCATGCGCCAGCACCTGCGCGCCCTGGGCCGG 77000
QY 1826 -----lletp-----ArgProGlyThrGluGlnSerSerG1 1836
Db 76999 GAACACGAATGCCGTCTTGCCACCGCTCGGCGCGCCGCCACACAGTCGCGCGCGG 76940
QY 1836 ySerSerGlySerSerGlyGlyGlySerSerSerArgProAlaSerHisSerHi 1856
Db 76939 CACG-----CCGCGCGCAGACCGCTGACGCC-TTCCA 76908
QY 1856 s-----AlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
Db 76907 TCAGTCAGTGGCGTGGCGGCCACGACACGCGCGCGATG----- 76868
QY 1870 aLeuGlnArgProSerValLeu-----HisAsnThrGlyMetLysG1 1885
Db 76867 -ATCAACACGCGCGCGGTGTGA-CCAGCGACACCCACATCGAGCGGCGCACGT-- 76812
QY 1885 ylleIleThrAlaValGluProSerLysProThr---ValLeuArgSerThrSerTh 1904
Db 76811 -----TCTCATCGCGCGCCGCCAGCGCCACAGCGCTGCGCGCTGGTTCAGCGCT 76759
QY 1904 rSerProValArgProAlaAla---ThrPheProAlaThrHisCysProLeuGlyG1 1923
Db 76758 TGGCGCACTCGCGGACAGACCCAGCGACCGCCACCGCGGTGCTTC----- 76710
QY 1923 yThrLeuAspGlyValTyProThrProValLeuMetGluProValLeuProLysGluAlaPr 1943
Db 76709 -----CATGGCGCGCGCAACCGGTTCCTACTGCGCGTGCCTCTCCAGAAATCACATGT 76657
QY 1943 oArgValAlaArg----- 1947
Db 76656 GCGTGTGTGCGCTGATCCCGAAGACGACACACACCGCGCGCGCGCGCGCGCAGC 76597
QY 1948 -----ProGluArgProArgAlaAs 1954
Db 76596 TGGCAGCGCGCGCTCGGTCAACACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 76537
QY 1954 pThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSe 1974
Db 76536 GTCGGCACATCCAGTCAACGTCTTGGCGAGCAGCGCGCGCATCGCTGCACCATC 76477
QY 1974 rProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrII 1994
Db 76476 T-----TGATCACCGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 76426
QY 1994 eAlaArgThrProAlaLysAsnLeu-----AlaProHisAlaSerProAspPr 2011
Db 76425 TTGATGACCCCGCCGACCAAGCGCGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 76366
QY 2011 oProAlaPro-----ProAlaSerAspProHisAr 2023
Db 76365 GCCTGGCGCTCAATGGGATCCCCAACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 76306
QY 2023 gGluLeuThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTy 2043
Db 76305 ACGTGCG-----CGACACCTAACCGC 76285
QY 2043 rHisGlySerSerTySer-----ProGluGlyValGluPr 2055
Db 76284 GCACTG-GCCAGCGCGCGCATGATACCCCTTGTGCGCGCGCGCGCGCGCGCGCGCGCG 76231
QY 2055 oValSerProValSerSerProSer---LeuThrHisAsp-----Ly 2068
Db 76230 -GCCAGCGCGTGGAGCGCGCGTCTCTGATTGACCGCGCGATCCCGCGCGCGCGCGCGCG 76172
QY 2068 sGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyLeuArgPr 2088

Db 76171 CGATCCCGCAACCGCGCGCGCTCAGCCAAACCGCGCCA--GCACGAGCACCCCGCGCGCC 76114
QY 2088 oLysGlnPro-GlyProVal-----LysLeuGlyGlyG 2099
Db 76113 TCGGCAACCGCGTCCATCGCGCGCGCACCGCATAGCCCTTACACGACCATCGCGGAC 76054
QY 2099 luAlaAlaHisLeuProHisLeuArgProLeuPro----- 2110
Db 76053 AG-----CGCCCGCTGCTGTAACATCGATGAACATCGCGCGG 76015
QY 2111 -----GluSerGlnProSerSerSerProLeuLeuGlnThrAlaPro----- 2124
Db 76014 GTGGCCATCAGCGTGACCCACCGACGACGCGCGCGATTCGCGAGGCGCAGTGAC 75955
QY 2125 -----GlyValLysGlyHisGlnA 2131
Db 75954 TGCACCGCCAAATGCGAGTCCCAACGACGACGACGCGGTATCCACGACACCGCC 75895
QY 2131 rgValValThrLeuAlaGlnHisIle----- 2139
Db 75894 GGGCCCTCGAG--GCCCAACACATACGCCACCGCGCGCGAGGCCACGCTCAGCGTCGAG 75838
QY 2140 --SerGluValIleThrGlnAspTyThrArgHis----- 2150
Db 75837 CCACGACGCGCGTAGCGTCCAGGTCCCGCGCACCGCGCTTGGGCCCCATACGAGCG 75778
QY 2151 -----HisPro-----GlnGlnLeuSerAlaProLeu-----ProA 2161
Db 75777 TGGAAACACCGCGCGAGACCGCGCTCTGCGAACCCCGACACGATGGGTGCGATGCC 75718
QY 2161 laProLeuTySerPhePro----- 2167
Db 75717 GCCCGCTCAACCGTTCACGACGACTTCCACAGCAACCGCTGCTGGGGATCCATCGCA 75658
QY 2168 -----GlyAlaSerCysProValLeuAspLeuArg 2177
Db 75657 AGCGCTCGTGGTGGTTCGCGTCCGAAAACTCGCGCTCAAAACCGCGGACATCCGTCAG 75598
QY 2178 ArgProProSerAspLeuTy---LeuPro-Pro-----Pr 2188
Db 75597 AACCGCGCGCACCGGTATACGACTTGCCTACCGCTCGGGATCGGGGTCAACACCC 75538
QY 2188 oAspHisGlyAlaProAlaArg-----GlySerProHisSerGluGlyLysAr 2205
Db 75537 GCCACAT-----CCAGCGCGCATCGCGCGGAACTCCGAAACCGCGTCACGCGCGTGT 75484
QY 2205 gSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProva 2225
Db 75483 GCCACGAGTCCACACGCTCTCCGCGGAATCCACCCCGGATAGCGGATCCCATC 75424
QY 2225 lSerProPro-----GluGlyMetThrGluProGlyHisSerArgSerAl 2240
Db 75423 CCACACCGCGCACCGCTCGGTGGTCCGGGCAAGAGATCCCGTTC-TCTCGCTCAG 75365
QY 2240 aValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySerLy 2260
Db 75364 T-----CGCTCGTTTGTGAGCGATCTCGCGAGGCGCTTACGAG 75323
QY 2260 sSerProGly-----AsnThrSerGlnProProAlaPhePheSer---LysLeuTh 2276
Db 75322 GTCTTCGGAGTGTCTTTCATGCGAGGTTCGCGCGAGCGTAACTGCTGCGAGATTG 75263
QY 2276 rGluSerAsnSerAlaMet---ValLysSerLysLys-GlnGluIleAsnLysLysLeuA 2295
Db 75262 CGATCCGATAGCGAGTGGCGTTCGCTCGCGGAATCCAGCATCAGTGCAGCGCAGCA 75203
QY 2295 snThrHisAsnArgAsnGluPro---GluTyArgAsnLysSerGlnProGlyThrGluIleP 2314
Db 75202 ACACGAGTTCGGAACAGAAAGCGCTGGGCCCATGCGCATCATCAACCGCGGACTTCCGCTG- 75144
QY 2314 heAsnMetProAlaIleThrGlyLeu---MetThrTyArgSer----- 2329

Db 78652 CGGCACACCCCTCAGGAGAGTGACGTCGGGAAACACACCTCGGCCCCCGCCGCCACAAGGC 78593
Qy 1105 roProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaI 1125
Db 78592 CT----- 78591
Qy 1125 leSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProV 1145
Db 78590 -----GCAAAACCCCGAAACCGCCGCCCATACCCATACCCCGCCGCCG- 78549
Qy 1145 alGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheS 1165
Db 78548 -----CCAACACT--GATACCGTCGGCGAGCTCCA--T 78518
Qy 1165 erGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuG 1185
Db 78517 CGCCCGTCCCAACCGCGGCCACACCG-----ACAGCTCCGCGCGCGCTGCACCGA 78464
Qy 1185 lyValProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeu-----GlySerV 1203
Db 78463 GCCAACCCCAACACGCGCTCG-----CGTGCAACCTCCATCCCGGACTCGCAATC 78413
Qy 1203 alProGly-GlySer---IleThrLysGlyIleProSerThrArgValProSerAspSer 1221
Db 78412 CGCCTGGCGCGGACCGGATACACCCACATTCGAGTACCGACTG-----CTCGACGCC 78356
Qy 1222 AlaleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGly 1241
Db 78355 GCCCACCACCACT--GACCGCGCACCCCAT---CTGCGCGGACGACCAACCGTCCCTGCA 78300
Qy 1242 ThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAsp 1261
Db 78299 ACCTCAACTCCT--GCAGACCGGAACCAACCCCTCTGCTACCGCGCCGCAAGGCCA--- 78246
Qy 1262 SerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGlu 1281
Db 78246 ----- 78246
Qy 1282 GlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProPro 1301
Db 78245 -----ACTCAAAACCCCGTGC CGCGGAACAGCACCACCGCGGCCACCG 78201
Qy 1302 HisGluThrAlaAla-----ProLysArgThrTyrAspMetMetGluGlyArg 1317
Db 78200 CATGATCGCGCGACCGCGGACGACCGATATCCGCGCCGCTCAACACCCACCCAC 78141
Qy 1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db 78140 CCGAGTCGGCGCTT-----GCACCACTGCAGCCCAACAATCCATGCTCGGCGCCACCA 78087
Qy 1338 GluArgHisSerProHisLysLysGlnGlnHisIleArgGlySerIleThrGln 1357
Db 78086 AGCCGACACCAACCCAGATCGC---CGCCACAGCACCTAGGCGGCTAGCCAGAACCGCT 78030
Qy 1358 GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgGlyAlaLysLeu 1377
Db 78029 GGC-----CGCA-- 78021
Qy 1378 LeuLysArgGluGlyThrProProProProProProProProProProProProProPro 1397
Db 78020 ACGCGTAGTCGGCAACTCCACCGCGCGCTCCAGGTCCGCAACCGCGCACCGCCAGT 77961
Qy 1397 rlySerGlnAla-----LeuGlyProLeuLysLeuLysProAlaHisGluG 1413
Db 77960 CAACACCCACCGCCGACAGTGGGCTGCCCGCGCAGCAGCCAAACCGGCCACCCCGC 77901
Qy 1413 lyValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLe 1433
Db 77900 GTCATCGGACCCAGCT-----CGGAATGACAA 77871
Qy 1433 uArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrG 1453
Db 77870 TCGGTTCCAC-----CGTAGCGCGCGATCACAGT 77841

Qy 1453 nGlyThrProLeuLysTyrAspThrGlyAlaSerThr-----ThrGlySerLysLy 1470
Db 77840 CGACCAACAGTCTC---TTCGATCCCGCGGATCAGGACCGATGGGGCTGGATTCGACGA 77784
Qy 1470 sHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLe 1490
Db 77783 ACACCCGTCGCGCTCGAAGCGTTCGGGACGGCGCGCTTC-----GAACCT 77736
Qy 1490 uAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSe 1510
Db 77735 GCACCGCTCGCGATGCT----- 77717
Qy 1510 rArgProGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleVa 1530
Db 77716 -----TCGTACACGACTACTCGG 77700
Qy 1530 lProGluLeuGlyLysProArgGln-SerProLeuThrTyrGluAspHisGlyAlaProP 1550
Db 77699 CGTTCACACCGCGGTATCCATGAGTTCCGCGGTGACAGTGGAG----- 77656
Qy 1550 heAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArg--- 1568
Db 77655 -----AAGAACGCCACCGTGAAGTACGGGTTCGATCTCGATCGTCCGACGC 77613
Qy 1569 -----LeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrS 1587
Db 77612 CGCGGATGAGCTCTCCGCGATCGCTCCACTCGCGCGAGTGTGACGCGTAGTCGACGT 77553
Qy 1587 erThrProArgGluIleAlaLysSer-ProHis-----SerThrValProGluHisHis 1604
Db 77552 CGATCCTCGCGGACGAAATGCTTCGGCTTCACATCGCTGCATCGATCGTCCGTCACGCGAT 77493
Qy 1605 Pro-----HisProLleSerProTyrGluHis-----LeuLeuArgGlyValSerGly 1620
Db 77492 CCCTCTCGCGCGCAGCACCGACGACGACCAATTCGATGTCGACGATATTCAGTCGCT 77433
Qy 1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
Db 77432 -----CTCCCATTTGGG 77421
Qy 1641 IleProLeuAspAlaAlaAlaTyrTyrLeuPro-ArgHisLeuAlaProAsnProTh 1660
Db 77420 ACGCAACTTCTCGCGCTGGCGTGGCCACAGGCGCAACGAGACCATCGCGCGCACCGC 77361
Qy 1660 rTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuG 1680
Db 77360 CCAACCCGACCGCAACCG----- 77342
Qy 1680 uAsnArgGlnThrIleAsnAspTyrIleThrSerGlnGlnMethHisHisAsnThrAl 1700
Db 77341 -----GCTGCGCAGTGCCACCCAGCAGCGC 77316
Qy 1700 aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLe 1720
Db 77315 CGTCCCAAGCGACAGGCGCGC-----CGCCACATATGCGCGCGCAT-----CT 77271
Qy 1720 uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleAlaAspLeuSerGlnValProHisLe 1740
Db 77270 CCCCCTCGGAAT--GCCCGATGACCGCGT-----CGGGNACCACACCCC 77230
Qy 1740 uPro-----ValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLe 1758
Db 77229 ACCGACCGCCACAATTCGGCCAGCGACACCATCCCGCCCAACAG----- 77184
Qy 1758 uAlaTyrLeuProThrAlaProGlnProPheSerSerSerArgHisSerSerSerProLeuSe 1778
Db 77183 -----CCGCTGCACCATCCACC--CGATCC-----AG 77156
Qy 1778 rProGlyGlyProThrHisLeuThrLysProThrThrThrThrSerSerSerGluArgGluAr 1798
Db 77155 CCGCGTTCGCGC-----GGTGGCCCGCGCAGCAGCTCGAGCAGCAGCAGCAGCTCC-- 77107

QY 455 aserPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLy 475
DB 80365 -----CTGCCACCGCGGACCATCGGGGATG-----CAACACGA 80331
QY 475 sAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnG 495
DB 80330 CTTTGGGATATGACGGGCTGACTGACAAACCGGTAGGCGCGCGGCGCATCGGCGAT 80271
QY 495 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 514
DB 80270 CAAAAGCTTGTACCGCAACCGCGCAACACTCCGCGCGGTAGCAACC-CACTACTCTG 80212
QY 514 gSerSerGlnGluGluLysAspGluLysGluLysGluLysGluLysGluLysGluG 534
DB 80211 GAAAGCA----- 80205
QY 534 uLysProGluValGluAsnAspLysGluAspLeuLysGluLysGluLysGluLysGlu 554
DB 80204 -TGCCTGCAATGCGGTCTGGGCGG-----CTTGGACCAATCGAACGCGCGCATACCG 80152
QY 554 rGlyGluAsnAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGlu 574
DB 80151 ACCCCCGATCGCGCTCGGCCACCGTCTGCGCGGATATCGGTT-TTACCCAGCTC 80093
QY 574 rGlnGlyArgArgLysGlyArgLysGlyArgLysGlyArgLysGlyArgLysGlyArg 592
DB 80092 GATAAGCGGCCACCGCTGGCGAGTAGCGCAGCGAGCATCGTGAACCTCACCGCGGAG 80033
QY 592 rGluGluAlaThrProGlnGlnSerAlaGluLeuAlaSerMetLeuLeuAsnGluSe 612
DB 80032 CGAGTTCAGCAGCTACGTCACACCGTGGCTGGTGGCGCGGAGAAACCGCTCTCTCGAA 79973
QY 612 rSerArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
DB 79972 CTCAACGATCGGAGTCGAGATATGATATCGTCAAAACCCATC-----GCCCG 79922
QY 632 gAsnTrpSerAlaAlaArgMetValGlySerLysThr----- 645
DB 79921 CAATGTATCCACTTGGCGGAGTCTGCGGTGACGAAACCTCTGCACCCCAATACGCGC 79862
QY 646 -----ValSerGlnCysLysAsnPheTyrPheAs 655
DB 79861 CAACGACACCGTCGCATACCAACCGCGGTGCGCGCATGACCAACACCTTC----- 79807
QY 655 nTyrLysLysArgGlnAsnLeuAspGluLeuLeuGlnGlnHisLysLysMetGluLy 675
DB 79807 ----- 79807
QY 675 sGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaAlaSerGluAlaAl 695
DB 79806 -----TGCCCGCGCGGACCTCTGCCCAACACCGA 79778
QY 695 aPheProProValValGluAspGluMetGluAlaSerGlyValSerGlyAsnGluG 715
DB 79777 CAGCCCG-----TAAACCGCGTCAAGAACCGCCCGGACCGCGCT----- 79735
QY 715 uGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyG 735
DB 79734 -----GCCTCCACCAACGACGAGCC----- 79714
QY 735 uCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHi 755
DB 79713 -----GCCCGCACCATGTTTACCAACCGCATCACCGCGCTCCGAACCC----- 79666
QY 755 sThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAl 775
DB 79665 -----ACCAGCCCAATAACCCCAT-CACCCGCTCACCAACGCGCA 79626
QY 775 aAspGlyProProGly-----ProProThrPro-----Pr 786
DB 79625 AACCGGTTA-CCCCCGGCGGACTTCAACGACCAACCGCTGCCCGTCGACACCGAGTTC 79567
QY 786 oArgArgThrSerArgAlaProIleGluProThr-----ProAlaSe 800

DB 79566 CGCCCGCGGGATATCCCAACGCAACCAACATCCCGAAATTGACCCCAACCGCA 79507
QY 800 rGluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProProPr 820
DB 79506 CCACACCGCACCGGACCTGCGCGCTGCGCAGTCTCACCGGACCAACGCGGCGCACACC 79447
QY 820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGlu---G 839
DB 79446 ACATCCGCCAAAGTCCCGCCACCAACGCGCAACCAACGCGCAG-CCCCCAGCGACCGCA 79388
QY 839 uGlyGlu----- 841
DB 79387 AGCNAAGTCAACCCGCGCGCTACCGGGGCAACCGTGGCGCATATACCGGGCTCGACG 79328
QY 842 -----GluGlnLysPro---ProAlaAlaGluGluLeuAlaVa 853
DB 79327 AACCAACGCTGGGCTCACCCGAGGTGGCGGCAACCAACCGCATCTCTCGACAGGAC 79268
QY 853 lAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGlu--- 872
DB 79267 CGACCATCGGTATCCACCAACCAACCGCGCGCGGTGTTACGCTGGCGCAGAGCGAC 79208
QY 873 -----GlyProAlaLysGlyLysAspAlaGluAlaGluAlaThrAl 887
DB 79207 CAACCCCAACCGCGAGCACCGCGCAATCGGTGACATCGTCAACCGGACCGCACCGC 79148
QY 887 aGluGlyAlaLeuLysAlaGluLysGluGlyGlySerGlyArgAlaThrAlaLy 907
DB 79147 CCA-----CGGTCAACCAACCAACC 79127
QY 907 s-----SerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAs 924
DB 79126 CAGACCGGTGCGATCGCTAGCAACCAACGACGCTCAACACCTCCACACCTCGTGACACC 79067
QY 924 pGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro-ArgProSer- 943
DB 79066 CCGGTACACCGCTGCCAGC-----ACATCCCGCGCGCGCTGCGC 79028
QY 944 -----LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysP 960
DB 79027 GCGCGGTGAGCGGCGGAGCTCCACA-----CCAGCGGTGCTGCCAATGTGCGCGCC 78974
QY 960 roLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaPro---ProIleGlnValT 979
DB 78973 CT---CAAACGCGCACCGGTGTCAGATCATCTCCAGCAATCCCGGTGCGCGCGGTGCA 78917
QY 979 hrlYsValHisGluProProArgGluAsp-----AlaAlaProThrLysProA 995
DB 78916 CTTGCAACGCGCGCGCTGAGACCGGCAACCATCAACTGCCGCGCACCGCAACACCGG 78857
QY 995 laProProAla-----ProProProGlnAsnLeuGlnProGluS 1009
DB 78856 CAACCTGCGGATCGCCCACTCCACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78797
QY 1009 etAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProp 1029
DB 78796 CACACGACCGCGCGGAGCGC-----CGTTGCGGTGCAACACACACACCC 78755
QY 1029 roAlaAspLysGluAla-----PheAlaAlaGluAlaGlnLysLeuProGlyAspProp 1047
DB 78754 CTGCCACGAGAACCGCAACATCGTCTGCTGCTCACCTCGACAAATTCGCC----- 78705
QY 1047 roCysTrpThrSerGlyLeuProPheProValProArgGluValIleLysAlaSerP 1067
DB 78704 -----ACGCATGCAACCGCGCATCCCAACCGCGCGGATGAA----- 78669
QY 1067 roHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuG 1087
DB 78668 -----TCCCAAAACCCCGTAT 78653
QY 1087 lyLeuHisAspThrAla-----ArgProValLeuProArgProProThrIleSerAsnP 1105

| | | | | | | | |
|----|-------------|---------|---------|----|-------|---|-------|
| WP | AAI99682_06 | 600001 | 710000 | QY | 137 | LysAspArgSerLeuThrGlu | 143 |
| WP | AAI99682_07 | 700001 | 810000 | Db | 81134 | ---GACCGGCGCAACTGGCTGTAATATGCGGTAGCGCAGCAGCAGCGCGCGCCAGCG | 81078 |
| WP | AAI99682_08 | 800001 | 910000 | QY | 144 | LysLeuGluProValSerProProSerProHisThrAspProGluLeuGlu | 161 |
| WP | AAI99682_09 | 900001 | 1010000 | Db | 81077 | CGGCACGATCCAGCGCGCGCCACTACACCGCGCATCGCTGCAGCAGCGTGTCCA | 81018 |
| WP | AAI99682_10 | 1000001 | 1110000 | QY | 162 | ---LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgVal | 179 |
| WP | AAI99682_11 | 1100001 | 1210000 | Db | 81017 | CGAACCTAGCGCTGCTCGG | 80997 |
| WP | AAI99682_12 | 1200001 | 1310000 | QY | 180 | AspArgGluIleThrMetValGluGlnIleSerLysLeuLysLysGlnGlnGln | 199 |
| WP | AAI99682_13 | 1300001 | 1410000 | Db | 80996 | ---TGGTTAGCGGAGCGCGCGCTGCGTCATCTCGCGCGATGCC | 80952 |
| WP | AAI99682_14 | 1400001 | 1510000 | QY | 200 | LeuGluGluGluAlaLysProGluProGluLysProVal---SerProProPro | 218 |
| WP | AAI99682_15 | 1500001 | 1610000 | Db | 80951 | GCTCGCGGAGTGGCGGGTTCATCGCGAGCGCTGCCACAGTCCCGCAGCCGAC | 80892 |
| WP | AAI99682_16 | 1600001 | 1710000 | QY | 219 | IleGluSerLysHis-ArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLeu | 238 |
| WP | AAI99682_17 | 1700001 | 1810000 | Db | 80891 | ATCGCGCCAGCCCGCGAGCG | 80870 |
| WP | AAI99682_18 | 1800001 | 1910000 | QY | 238 | aGluAlaAlaHisArgIleLeuGluGluLeuGlyProGlnValGluLeuProLeuTyrAs | 258 |
| WP | AAI99682_19 | 1900001 | 2010000 | Db | 80869 | ---CGATAGGCCA | 80859 |
| WP | AAI99682_20 | 2000001 | 2110000 | QY | 258 | nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLys | 278 |
| WP | AAI99682_21 | 2100001 | 2210000 | Db | 80858 | CCAACCC | 80832 |
| WP | AAI99682_22 | 2200001 | 2310000 | QY | 278 | sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh | 298 |
| WP | AAI99682_23 | 2300001 | 2410000 | Db | 80831 | CAGCATAAT | 80822 |
| WP | AAI99682_24 | 2400001 | 2510000 | QY | 298 | eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAs | 318 |
| WP | AAI99682_25 | 2500001 | 2610000 | Db | 80821 | ---CCCTGAGCGCGGTGCGCCA | 80802 |
| WP | AAI99682_26 | 2600001 | 2710000 | QY | 318 | nAsnProArgArgAlaLysGluSerLysValArgGluTyrTyrGlu | 334 |
| WP | AAI99682_27 | 2700001 | 2810000 | Db | 80801 | CAATCCCGCATCGACGAAACACACCAACCCGCAAAATCCATGCTCTCGGTAGCT | 80742 |
| WP | AAI99682_28 | 2800001 | 2910000 | QY | 335 | -LysGlnPhePro-----GluIleArgLysGlnArgGluLeuGlnGlu---ArgMe | 350 |
| WP | AAI99682_29 | 2900001 | 3010000 | Db | 80741 | CGTCAGATTCCAGGCGCGCATCGACCTTGGCCGCAACACCGTATCCACCGATCCGGTG | 80682 |
| WP | AAI99682_30 | 3000001 | 3110000 | QY | 350 | tGlnSerArg-----ValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaAr | 367 |
| WP | AAI99682_31 | 3100001 | 3210000 | Db | 80681 | TCAGCCCGTATCAGCGCATCGTCCAAACACCCCGCGCGCATG | 80639 |
| WP | AAI99682_32 | 3200001 | 3310000 | QY | 367 | gSerGluHisGluValSerGluIleLeuAspGlyLeuSerGluGlnGlnGluAsnLeuGlu | 387 |
| WP | AAI99682_33 | 3300001 | 3410000 | Db | 80638 | ---AAACACCCCTTTAAGCGGATA-----GCGCGGATCCAGACCCGCCA | 80598 |
| WP | AAI99682_34 | 3400001 | 3510000 | QY | 387 | sGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp----- | 401 |
| WP | AAI99682_35 | 3500001 | 3610000 | Db | 80597 | ACAATGCGCGCAGCGCATCAGCATCAGCAGGAGACACACCGCCACCTGGGCCCC | 80538 |
| WP | AAI99682_36 | 3600001 | 3710000 | QY | 402 | ---AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMet-----Al | 417 |
| WP | AAI99682_37 | 3700001 | 3810000 | Db | 80537 | CGCCCTCGCGCAACAGGGCGCGGACTTCTGCGCGCCCTGTGCGCTGTCTCCCGCTTCGGC | 80478 |
| WP | AAI99682_38 | 3800001 | 3910000 | QY | 417 | aAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlu | 437 |
| WP | AAI99682_39 | 3900001 | 4010000 | Db | 80477 | TGACAGAACACAGATGGCCACCCCATGTCGCCGACCAATGGGTAGTACCCCGCGAAC | 80418 |
| WP | AAI99682_40 | 4000001 | 4110000 | QY | 437 | sGluThrPheArgGluLysPheMetGlnHis-----ProLysAsnPheGlyLeuIleAl | 455 |
| WP | AAI99682_41 | 4100001 | 4210000 | Db | 80417 | CGGCCATCGCGGTCCCCCGGATGACACCGGTGCGCGCGCGCGCGCGCGCGCGCGCG | 80366 |
| WP | AAI99682_42 | 4200001 | 4310000 | | | | |
| WP | AAI99682_43 | 4300001 | 4410000 | | | | |
| WP | AAI99682_44 | 4400001 | 4411529 | | | | |

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

1,43e-08

581.50

31.25%

22.48%

4.40%

4

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

110000

654

255

954

1060

150

US-09-522-753-5 (1-2517) x AAI99682_18 (1-110000)

| | | | |
|----|-------|--|-------|
| QY | 58 | IleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsn--- | 76 |
| Db | 81412 | GTAGTCGAAGATCAGCGTGGCGA-----CAACGTGAGCGCGCGTGGCGGT | 81368 |
| QY | 77 | ---GluArgSerGlnGluLeuHis-----LeuArgProGlu----- | 87 |
| Db | 81367 | CTTGAGTCGTTGCGCGAGTCTCCAGCGCGTCAAGGAATCGAACCGAGATCTTGAATGC | 81308 |
| QY | 88 | -----Ser | 88 |
| Db | 81307 | TTTGTGAGCGTTGATATCGACACTGCTGGAACACCAACCGCGCGCGCATGCTGAT | 81248 |
| QY | 89 | HisSerTyr-----LeuProGluLeuGlyLysSerGluMetGluPheIleGlu | 104 |
| Db | 81247 | CACCAATCGGTGAGTTCGGCGTCCCGCGCTCGGCGGTGAGCGCATGCGA----- | 81197 |
| QY | 105 | SerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeu | 124 |
| Db | 81196 | -----CCGGCTGA-CTAAGC | 81183 |
| QY | 125 | AlaThrGlyGlnPro-----AlaGlySerGluAspLeuThr | 136 |
| Db | 81182 | CCGACATCGACACCGCGGTATCGCGCGGTGTCGATCTCTCGCGCGGTGCG | 81135 |

| | | | | | |
|----|------|---------------------|--|---|------|
| Db | 8007 | TGACCCCGTCA | CCCTTAAATCGAGACCAAGGTC | TCTTCAGCCGCGCAAACTGGGGTCCAC | 8066 |
| Qy | 2001 | nLeuAlaProHis | HisAlaSerProAspProAlaProAlaSerAlaSerAsp | Pr | 2021 |
| Db | 8067 | GCTCAGCC | CCCCAC | -----CTCTCTGCTCTGCCAGCAAACTGCTCTCCATTTGGC | 8114 |
| Qy | 2021 | o-----His | ArgGluLysThrGlnSerLysProPhe-----Ser | IeGlnGluLeu | 2037 |
| Db | 8115 | AGTCAACCATG | TCCCTCGGGCCAGCATCC | CCAGCAGATCGAACTGTCTCCCATTTGGC | 8174 |
| Qy | 2037 | uLeuArgSerLeu | GlyTyrHisGlySerSerTyrSerPro-GluGlyValGluProVal | S | 2057 |
| Db | 8175 | AGCTGCAAA | GCTAGATGCTCAT-----TCTCTCGAACCAAGTCGACCCGGGCC | 8222 | |
| Qy | 2057 | erProValSer-----Ser | ProSerLeuThrHisAspLysGlyLeuProLysH | 2073 | |
| Db | 8223 | ATCCTCATTTCC | AAGGCAACCCAGCAGTACTGCAT----- | 8263 | |
| Qy | 2073 | isLeuGluGluLeu | AspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly | 2093 | |
| Db | 8264 | ----- | -----CTACGGCGCTCTCCACCAACGC | 8285 | |
| Qy | 2093 | roValLysLeu | GlyGlyGluAlaAlaHisLeuProHisLeuArgPro----- | 2108 | |
| Db | 8286 | CACAGTCATG | TGGCTCAGGCATCC | CAGTGGCCCGCTTCATCTCCAGCATCCACCCAGA | 8345 |
| Qy | 2109 | -----Leu | ProGluSerGlnProSerSerProLeuLeu----- | 2120 | |
| Db | 8346 | GCAGTCTGT | CATCATGTCACCCACAGCATCACCAGACTGTCTCCCTGAGCCACCTCTC | 8405 | |
| Qy | 2121 | ----- | -----GlnThrAlaProGlyValLysGlyH | 2129 | |
| Db | 8406 | CCAGGGCGAG | TGAGATGAACACTCC | CCCGTCCAGTACACCTACAGCATCCGGCC | 8465 |
| Qy | 2129 | isGlnArgVal | ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr | 2149 | |
| Db | 8466 | AGAAGCGCTT | CACCTCTCTCGGGCTCGCTGCAGCCC-----CAGCAAAATAGAGG | 8515 | |
| Qy | 2149 | rgHisHisPro | GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyA | 2169 | |
| Db | 8516 | TCAGGGCC | CCACAGCGCCAGCACCCGAGCCGAGCCGCCA-----G | 8557 | |
| Qy | 2169 | laSerCysPro | ValLeuAspLeuArgArgProProSerAsp-----LeuTyrLeuP | 2186 | |
| Db | 8558 | CTGGTGTG | CTGCACCTGCGCTCC | CCAGCACCTCCCGAGGAGGAAGTCAATTATCACCTTC | 8617 |
| Qy | 2186 | roProProAsp | HisGlyAlaProAlaArgGlySer----- | 2197 | |
| Db | 8618 | CTGTGCTCG | AGCCACAGCCCTCTGCAGTCAGAGGTACTAGTCATGCATCTCGAGTACC | 8677 | |
| Qy | 2198 | ----- | -----ProHis----- | 2199 | |
| Db | 8678 | GACTGCAC | CCCTATACTGTGCCACGGGATGTAGGATCATGTGGTCATCCACATGTGACGG | 8737 | |
| Qy | 2200 | ----- | -----SerGluGlyLysArgSerProGluProAsnLysT | 2212 | |
| Db | 8738 | CAGTCAGCAG | CACCCCGGCCGCGGTGGGTGAAGTGCACACAGCCAGCAAGG | 8797 | |
| Qy | 2212 | hrSerValLeu | GlyGlyGlyAlaAspGlyIleGluProValSerProProGluGlyMetT | 2232 | |
| Db | 8798 | CCCCCT----- | -----CAGCAGCCAGGAAGGAGTGC | CAACACACAGATGCCAAAG | 8845 |
| Qy | 2232 | hrGluProGly | HisSerArgSerAlaVal-----TyrProLeuL | 2245 | |
| Db | 8846 | CTGCCCCC | ACCCCTCCCTGCTCCCTGTCCCTGTCCCTTCCTGCCCCCTGCTC | 8905 | |
| Qy | 2245 | euTyrArgAsp | GlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnT | 2265 | |
| Db | 8906 | CTGCCCC | CTCATGGTGAG-----GCCGTATCTCTCACAGTTACCCCGAGTACC | 8953 | |
| Qy | 2265 | hrSerGln----- | -----ProProAlaPhePheSerLysLeuThrGluSerAsnS | 2280 | |

RESULT. 99.

AAI99682_18/c
Continuation (19 of 45) of AAI99682 from base 1800001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Sequence split into 45 fragments

| WP | Sequence split into 43 fragments | LOCUS | AAI99682 | ACCESSION | AAI99682 |
|----|----------------------------------|--------|----------|-----------|----------|
| WP | Fragment Name | Begin | End | | |
| WP | AAI99682_00 | 1 | 110000 | | |
| WP | AAI99682_01 | 100001 | 210000 | | |
| WP | AAI99682_02 | 200001 | 310000 | | |
| WP | AAI99682_03 | 300001 | 410000 | | |
| WP | AAI99682_04 | 400001 | 510000 | | |
| WP | AAI99682_05 | 500001 | 610000 | | |

| | | | | | | | |
|----|------|---|------|----|------|--|------|
| Qy | 1403 | GlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGly | 1422 | Qy | 1728 | oArgGlyIleAspLeuSerGlnValProHisLeuProValLeuValProThrPr | 1748 |
| Db | 6351 | ----- | 6351 | Db | 7086 | CACCTGGCTGGTGAACGCTCCCTGGTCCCG--GTGAATGCCCTGAAAGCCCTGTAA | 7142 |
| Qy | 1423 | ArgSerIleHisGluIleProArgGluGluLeuArgHisThr | 1437 | Qy | 1748 | oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPh | 1768 |
| Db | 6352 | AATTCTGTCCTCAAGAAATCAGTGTGAGGAAGAGACTCCAAACCAAGCATCTGTGCCCCCA | 6411 | Db | 7143 | GGGCTCAGTGACACACTGAAAGATTGGTGAGC-- | 7176 |
| Qy | 1438 | GluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeu | 1457 | Qy | 1768 | eSerArgHisSerSerProLeuSerProGlyProThrHisLeuThrLysPr | 1788 |
| Db | 6412 | GACCTTCCCCCACTCCCGACCA-- | 6444 | Db | 7177 | -----ACCCTGCTGGCGCCGTGAACGCTCTCTGAAAGG | 7208 |
| Qy | 1458 | LysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisaspValArgSerLeuIle | 1477 | Qy | 1788 | oThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspAr | 1808 |
| Db | 6445 | GATGAGGAGCTCAAGCC-- | 6486 | Db | 7209 | GCCTGTG-- | 7215 |
| Qy | 1478 | GlySerProGlyArgThrPhePro-- | 1495 | Qy | 1808 | gGluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaProIle | 1826 |
| Db | 6487 | GAAGTGACCCGGTGACCCACCCAGCGATCCCAAGCATCCCATACCCACTGCTCTCT | 6546 | Db | 7216 | -----AATGTTCTTACGGGGCCAGTGAATGTTCTCACCACCTCAGTGAACGC | 7262 |
| Qy | 1496 | AlaArgAlaLeuGluArgAlaCysTyrGluLysSerLeuLysSerArgProGlyThrAla | 1515 | Qy | 1827 | uAspArgProGlyThrGluLysSerSerGlySer | 1837 |
| Db | 6547 | GTA-- | 6555 | Db | 7263 | CACGGTGGGCACAGTGAATGCCGCCCGAGGCACAGTCAATGCCCTCGAGTGCAGTGAA | 7322 |
| Qy | 1516 | SerSerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLys | 1535 | Qy | 1838 | -----SerGlySerSerGlyGlyGly | 1844 |
| Db | 6556 | GCAAGCTCTCACTCTCTGCTCGCTCTGGG-- | 6591 | Db | 7323 | TGCCACACCAAGTCAGTCACGTCACAGCGGTGCGGTACTCTCATCTGCTGTGT | 7382 |
| Qy | 1536 | ProArgGlnSerPro-- | 1548 | Qy | 1844 | yGlyGlySer-- | 1851 |
| Db | 6592 | CCACACAGAGCCCCCTACTAAGGTGCACAGAGTGGATCAAGAGCAGGAG-- | 6642 | Db | 7383 | AACGCCCAACACAGCAGCTGACAAATGGCAGGGGCAGTGAATGCCCGTCAACAAGATG | 7442 |
| Qy | 1549 | ProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArg | 1568 | Qy | 1851 | oAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLe | 1871 |
| Db | 6643 | -----GAGCCAGGGCTCAGTCTACTCTCCAGCT | 6675 | Db | 7443 | CAAAACAGAGAGCGTGTATATGAACAGTCGGTTCACCCAGCGTCCATCGCTGTGT | 7502 |
| Qy | 1569 | LeuGlnGluGlySerLeuSerSer-- | 1585 | Qy | 1871 | uGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGly | 1891 |
| Db | 6676 | CTTCCCCACAGACAAAGCGCTCTGATGTGACACCAAGCTCCAGCACCCTGAGGAAGATT | 6735 | Db | 7503 | CGACGATCGTCCGGCA-- | 7550 |
| Qy | 1586 | ThrSerThrProArgGluIleAlaLysSer-- | 1605 | Qy | 1891 | uProSerLysProThrValLeuArgSerThrSer-- | 1908 |
| Db | 6736 | CTCARGGACCCCAAGATATGG-- | 6789 | Db | 7551 | CACCTCTGAAGGGTGTGCTCTCAGTACTCAGGGCAGAAGACCGAAGCCACACGCG | 7610 |
| Qy | 1605 | oHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValaspLeuTyrAr | 1625 | Qy | 1908 | gProAlaAla-- | 1925 |
| Db | 6790 | ACAGCAATTGCAGACGCTG-- | 6808 | Db | 7611 | GATCAGCGCAAGATCAGCCAGATCCCCCGGCCAGT-- | 7652 |
| Qy | 1625 | gSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAl | 1645 | Qy | 1925 | uAspGlyValTyrProThrLeuMetGluProValLeuProLysGluAlaProArgVa | 1945 |
| Db | 6809 | -TCAGTGCTGCCCTTGCTC-- | 6864 | Db | 7653 | GGCATTGAATTTTACAGCATGCTCCAGTCCCAAGTCCCAAGTCAAACTGATCTGTGCACAGC | 7712 |
| Qy | 1645 | a-- | 1654 | Qy | 1945 | lAlaArgPro-- | 1961 |
| Db | 6865 | CTTTTAGAAGAAAACAGCACCTCCAGTGACAAACAACTCTGAGATACAAAGCCTCGAG | 6924 | Db | 7713 | ATCGCAGCTCCATCCAAAGGCCCTCAAGCTCTGAGGCTATGCAACGTCGCCACCCA | 7772 |
| Qy | 1654 | s-- | 1668 | Qy | 1962 | -----LysProProAlaArgSerGlyLeuGly | 1970 |
| Db | 6925 | GTGCTGGTAGCTGCACAAAGAAAGTGGTCCAGTCATGCTCCCAAAATTACCTCT | 6984 | Db | 7773 | TTCCAGCTTGGTACTGACGCCCCAGACATATATGCTCTCTGTGATTTCTGTGTGAA | 7832 |
| Qy | 1668 | rLeuIleAsGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAs | 1688 | Qy | 1970 | uProAlaSerSerProSerLysGlySerGluProArgProLeu-- | 1984 |
| Db | 6985 | GTTATTAGCCGGATG-- | 7035 | Db | 7833 | G-- | 7886 |
| Qy | 1688 | pTyrIleThrSerGlnMethHisAsnThrAlaThrAlaMetAlaGlnArgAlaAs | 1708 | Qy | 1985 | -----ValProProVa | 1988 |
| Db | 7036 | -----ACCTTGGCAAAACCGCTCCTCA | 7058 | Db | 7887 | GCCTGTCAACCCAGGAGGCACAGTGAAGTTCTCACCCAGGGATCAACACACCCCTGT | 7946 |
| Qy | 1708 | pMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyPr | 1728 | Qy | 1988 | lSerGlyHisAlaThrIleAlaArgThrProAla-- | 1999 |
| Db | 7059 | AACCTTCACTGGTCTG-- | 7085 | Db | 7947 | GCTGGTTCACACCGAGTGTCTCTCACCCTCCAGCATTTGTCAACCAACAAAGAGCTTGC | 8006 |
| Qy | | | | Qy | 2000 | -----LysAs | 2001 |

[illegible]

QY 390 -----ArgGlnLeuAlaValIleProPheMetLeuTyrAspAlaAspGlnGlnArg 406
DB 2281 AAAGTGGACAGACTTAATACGTGTGCCAGCCC-----AAAGACTGTCAGGAG 2328
QY 407 ILeIysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys----- 424
DB 2329 CTTGCCAGTATTCTGTGGTGGTCTGCTCAAGGCCAGCTCAGACCTACAGCAAGAC 2386
QY 425 ---AspArgGlnValMetAsnMetTyr-----SerGluGlnGlyLysGluThrPhe 440
DB 2387 TGGGAGAACTAGCAGGTGGAATCTGTGGAATAATCAAGAAGTCCCAATCAAAAAGCCCATTC 2446
QY 441 ArgGlnPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 2447 CCTCAAAACCCAGCTCAACAGAGCTGAGGTATATAGATGATCAAGGACCAAGAGAGAGAG 2506
QY 461 -----LysThrValAlaGluCysValLeuTyrTyr----- 471
DB 2507 ACCTTAGGAAAACCTATTGCCAGTCTCTGATGAACACACCTGAAACCTGTAATCAGGCCAAG 2566
QY 472 -----LeuThrLysLysAsnGluAsnTyrLysSerLeuValAlaArg 485
DB 2567 AGAAATCACATTCAGTAAATCTAGAGAAAATAATGCGATTCATGATCAGACGCGAGA 2626
QY 486 SerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 505
DB 2627 GT-TACCGAAAACAAATCGAACAGAGTCTGAGGAAACAGCAGATGGAATAGGAATAGCC 2685
QY 506 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 525
DB 2686 AAGTCTGAGAAGTTTGGCAGTCTCTAAAGAAAGATGTAGATGAATATGAAGACGTAGCCTC 2745
QY 526 GluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeu 545
DB 2746 GTTCAGAGGTAGGCAACCCCTCAA-----GATGTCACTGATGACTCTCTCCTAGC 2799
QY 546 LeuLysGluLysThrAspAsp-----ThrSerGlyGluAspAsn--- 558
DB 2800 AAAAGAAAAGGATGGATCATGTCGATTTGATATCTCCACCAAGCCGAGAAAGGAATAC 2859
QY 559 AspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArg 578
DB 2860 AGAAGTTCACGCCAAATCAGCGAAGATCTGNAAGACTGTGGTGTCTCCAGGTGTCGA 2919
QY 579 LysGlyArgIleThrArg-----SerMetAla 587
DB 2920 CATGGTCTCTCCATGACATGAGGATCCCATAGGCTCCCTAGGCTACTGTGAGTAAA 2979
QY 588 AsnGluAlaAsnSerGluGluAlaIleThrPro----- 598
DB 2980 GGGTCTCTCTAAAGTAGATGAAAAGTCTCTCCCTTATTTCAACATAACAGTCAGGGAAGAG 3039
QY 599 -----GlnGlnSerAlaGluLeuAlaSer 606
DB 3040 TCTTTAAATTTAATCCTTATGATTTCTACAGGAGAGAACAGATGGCAGATATGCCAAA 3099
QY 607 MetGluLeu-----AsnGluSerSerArgTyrThrGluGluMet 620
DB 3100 ATAAACTATCTCTTGAATTCGAAGATGAACATAATCGTTGGGACTCTCAGATGAAA 3159
QY 621 GluThrAla----- 623
DB 3160 CAGGATCTGGCAGATTTGATGTAGTTTCCCAACAGCATTAATTAAGAGATAGCCTT 3219
QY 623 ----- 623
DB 3220 CGAAAAGGTCTGTACGAGATCTGGAACCTGGTGTAGGTGCTTCTGATTTCTGACGAAGAT 3279
QY 623 ----- 623
DB 3280 GGTGAACACAAATCCCACTCACCAGAGCCTCTGCATTTATATGAAGATTCTCGATTGTCT 3339

623 ----- 623
3340 TTTTATTGAGGACAGAGAACAAGCTAGCTAGCGAGATGAAGACTCTCTAGTTCT 3399
624 -----LysLys 625
3400 TTAGAAAGAAACAAATTTTACTCTTTTTCATTGATTAAGACAATFCACACCAAGACTAAA 3459
626 GlyLeuLeuGluHisGlyArg-----AsnTyrSerAlaIle 637
3460 GCTTTGCTTGAAGAGCTTAATCTCTCTTCATCTCTGTAAGAAAATTTGGTCTTTCTT 3519
638 -----AlaArgMetValGlySerLysThrValSerGln----- 648
3520 GATTCGGGACTCCGATTTGCAATTTTTCGAAACAACAAGATGAAGAAAAGGTTGACTCT 3579
649 -----CysLysAsnPheTyrPheAsnTyrLysLys----- 658
3580 GCTCCAAGACCTATTTCATCTCTGTTACATGAAGAAAAGAAAATTTAGGACTGATTCAAA 3639
659 -----ArgGlnAsnLeu 662
3640 GGGAAAATGGATGATTAAGAAAGAGACCATTAAGAAGAAGAGAGAGAGAGAGAGAGATTG 3699
663 -----AspGluIleLeuGlnGln----- 668
3700 TTTGCTTCTCGTTTTTACACAGCTCAATCTTTTGAACAAGATTCACAGCGATTGCGCAT 3759
668 ----- 668
3760 CTAGAGAAAAGAGAGAGATTCGACTTCATTTCTGTTAGGATCTATGGGAAGCAGACA 3819
669 -----HisLys 670
3820 TCTGAGGAGCAACAGCACAATCTGATTCATTCAAGAACCCAGTAGTTCTGTTCCATAGC 3879
671 LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLys----- 685
3880 AGATTATGAGCTCACACGGATGCAACAGAAAAGAAAAGAAAAGAGAGAGAGAGAGAGAG 3939
686 -----AlaProAla 688
3940 GAGTTGAGAAAAGAGAGATCAGAGATCATCCCAAAACCCAGAAATCTGCTCTGAG 3999
689 AlaAlaSerGluGluAlaAlaPheProProValVal----- 700
4000 AATAAAGATTCAAGAACTGMAAACTCCACCTTCGTTGGGCTCCAAAGTGTCAAGTCGTA 4059
701 -----GluAspGluGluMetGluAla 707
4060 ACTCTAGAATCAGCCCCATCAGCAGTGAAGAACCACTGGTGACAAAACGGTAGAGCG 4119
708 SerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSer 727
4120 CTTTGTGTACA-----GAAGAGAGACTGTGGAGCAGCTACCTACCT-----GTCTCA 4164
728 GlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp 747
4165 GAAGAAGCAAAAGCTGCTGCTGAA-----CCTGCTCTGCTCCCTGTGGAAACAGCTGGAACA 4221
748 ThrGluSerIlePro-----SerProHisThrGluAlaAlaLys----- 760
4222 GTAGACCTGCCCCAGAGCAGACCCCGATGAAGAGCTGCCATGATGCTCGGGGTGTT 4281
761 AspThrGlyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProPro 780
4282 GAGGAGGTTTATCAGGTGACCCGCTCT-----TATCTGATGCTCAAG----- 4326
781 GlyProProThrProProArgArgThrSerArgAlaPro-----IleGluProThrPro 798
4327 ---CCTCAACTCCCGGGGCTCGTTTTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4383
799 AlaSerGluAlaThrGlyAlaProThrProProProAlaProProProSerProSerAlaPro 818

| | | | | | | | | | |
|----|-------|--|---|-------|-----------|--|---|-------|--|
| QY | 2092 | lyProVal | -----LysLeuGlyGlyGluAlaAlaHisL | 2103 | QY | 2330 | -----GlnAlaValGlnGluHisAlaSerThrAsnMetG | 2341 | |
| Db | 66978 | GTCCATCGCGGACCCGATAGGCTTACACAGCACCATCGCGGACAG | ----- | 66929 | Db | 65968 | ACAGATCCCGATTTCCGCCAACAGCGCCAGCTCAGCTCGAGAGCCTCGCGGCACAGGT | 65909 | |
| QY | 2103 | euProHisLeuArgProLeuPro | -----GluS | 2112 | QY | 2341 | lyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerP | 2361 | |
| Db | 66928 | -----CGCCGCTGCTGCTGAACATCGATGAACATCGCGGGTGGCCATCAG | 66880 | Db | 65908 | CGAGGGTCTCGGTGATGCGGTGATGATCTTGGGTCCC | -----CCGGGATGGGTGAGCCACAG | 65852 | |
| QY | 2112 | erGlnProSerSerProLeuLeuGlnThrAlaPro | ----- | 2124 | QY | 2361 | roProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSer | 2378 | |
| Db | 66879 | GTGACCCACCGACCGCCAGTTCGATTCGCCGAGGCGCAGTGACTGCACCGCCAAA | 66820 | Db | 65851 | CGCCAGCTCGGTG | -----GTGCTCAGCGCTCGGAAGCAGGAAGGTGTGACGTCATTGC | 65795 | |
| QY | 2125 | -----GlyValLysGlyHisGlnArgValThrL | 2135 | QY | 2378 | roAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProG | 2398 | | |
| Db | 66819 | TGCAGTCCACCAACAGACGACGACGACGCGGTATCCACCGACACCGCGGCGCTCGAG | 66761 | Db | 65794 | CCAGTACTGTCTGACACCGCGCGCAAGTCCCGT | -----GATAGGACGAGCTCGAACCCGG | 65738 | |
| QY | 2135 | euAlaGlnHisIle | -----SerGluValI | 2143 | QY | 2398 | ly | 2398 | |
| Db | 66760 | --GCCCAACATACGCCACCGCCGCGGAGGCCACGCTCAGGTTCGAGCCACGAGCCCG | 66703 | Db | 65737 | CCGAGCCGACGTCGTATCCCATGTCGCGAGCGAGTCCGGGTACAGATGGTGCAGGAT | 65678 | | |
| QY | 2143 | leThrGlnAspTyrThrArgHis | -----HisPro | 2152 | QY | 2399 | -----GlyGlyGlyLysAlaLysValSerGly | 2413 | |
| Db | 66702 | TAGCGTCCAGGTACACCGCGGACCGCGGCTTGGCCCTACGAGCCGTGGAACACCCCG | 66643 | Db | 65677 | CCAGGATGTCGCGCGCGCGCGATGCTCTGGGCGCGCTTACACCTGGCGCACCA | 65618 | | |
| QY | 2153 | -----GlnGlnLeuSerAlaProLeu | -----ProAlaProLeuTyrS | 2165 | QY | 2413 | ysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer | 2431 | |
| Db | 66642 | CGCAACACGCGCTCTCGGAACCGCGCAACGTCGATTTGGTTCGATCGCCCGCGCTCCAAC | 66583 | Db | 65617 | CGCGCGCGCGCTCAGCAAAACGCACTGCCACAGGCGCGGACGCTGCGGTGT | 65558 | | |
| QY | 2165 | erPhePro | ----- | 2167 | QY | 2432 | -----SerValHisSerGluGlyAspCysAsnArgArgThrProLeu | 2446 | |
| Db | 66582 | GCTTCCACGACACTTCCAAACAGCAACCGCTGTCGGGATCCATCGCAAGCGCTCGGTG | 66523 | Db | 65557 | ATCCCGGATACGTGAGTGAACACAGCTCGACCGAGACCAACCGCGGCGCGGCGG | 65498 | | |
| QY | 2168 | -----GlyAlaSerCysProValLeuAspLeuArgProProSer | 2181 | QY | 2447 | AsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPro | 2466 | | |
| Db | 66522 | GTCGCGATCCGAAACATCGCGCTCAAAACCGCGGACATCCGTCAAGAACCCCGCGCAC | 66463 | Db | 65497 | CCCCGCGCA | -----GGTAGTCTGTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 65441 | |
| QY | 2182 | AspLeuTyr | -----LeuPro | 2192 | QY | 2467 | LeuIleMetArgLeuGlnAlaGly | 2486 | |
| Db | 66462 | CGGTATACGACTTGCCTACCGCTCGGGATCGGGGTCAAAACACCCCGCACAT | 66408 | Db | 65440 | GCCGGAACAGCGCGCACCTCCGACATCGCGCGAGCCCGCGCGCGCGCGCGCGCG | 65381 | | |
| QY | 2192 | aproAlaArg | -----GlySerProHisSerGluGlyLysArgSerProGluPr | 2209 | QY | 2486 | a | 2496 | |
| Db | 66407 | CCACGCGCGATCCCGCGGAACTCCGAAACCGCGTACGCGCGCTGTCACAGCTCC | 66349 | Db | 65380 | CATCCAGCGAGCGCACCGCCAGTCCCGTGCCTGGCGGTGATCAACAC | 65332 | | |
| QY | 2209 | oAsnLysThrSerValLeuGlyGlyGlyLeuAspGlyLeuGluProValSerProPro | 2228 | Db | RESULT 98 | | | | |
| Db | 66348 | CACAGGCTCTCCGGGGAATCCACCCACCGGATAGCGGATCCCATCCCAACCCAGCC | 66289 | Db | AAC76700 | standard; cdna; 11680 BP. | | | |
| QY | 2229 | -----GluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLe | 2244 | Db | AC | AAC76700; | | | |
| Db | 66288 | ACCGGCTCGGTGTCGGGCAAGAGATCCCGTTC | 66241 | Db | XX | 08-FEB-2001 (first entry) | | | |
| QY | 2244 | uLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGly | 2263 | Db | XX | Human ORFX ORF2255 polynucleotide sequence SEQ ID NO:4509. | | | |
| Db | 66240 | -----CGCTCGTTTGTGAGCATCTCGGAGCGCTTTCAGGAGTCTTCGGAGT | 66188 | Db | XX | Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; | | | |
| QY | 2264 | -----AsnThrSerGlnProProAlaPhePheSer | 2280 | Db | XX | vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; | | | |
| Db | 66187 | GCTGTTTCATGGCAGGAGTCCCGGAGCGTAACGTCTCGGATGTCGATCCGGAATAG | 66128 | Db | XX | anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; | | | |
| QY | 2280 | rAlaMet | -----ValLysSerLysLys | 2299 | Db | XX | immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; | | |
| Db | 66127 | CGCAGTGGCGTTTACGTCGGGGAATCAGCATCAGTCGCGCAACACAGAGTTCGG | 66068 | Db | XX | hypotensive; dermatological; immunosuppressive; antirheumatic; antihypertensive; | | | |
| QY | 2299 | rgAsnGluPro | -----GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA | 2318 | Db | XX | antianemic; gene therapy; cancer; proliferative disorder; hypertension; | | |
| Db | 66067 | AACAGAGCTGGGCGCCATGGGATCATCAACCGGAGTTCGCTG | 66021 | Db | XX | neurodegenerative disorder; osteoarthritis; graft vs host disease; | | | |
| QY | 2318 | lalleThrGlyThrGlyLeu | -----MetThrTyrArgSer | 2329 | Db | XX | cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; | | |
| Db | 66020 | -----GGGGCGGTTTGGCGATGTGTCTACGCGACATCATGACGACGACGAGCTGACG | 65969 | Db | XX | cholesterol ester storage; systemic lupus erythematosus; infection; | | | |
| | | | | Db | XX | severe combined immunodeficiency; malaria; autoimmune disorder; asthma; | | | |
| | | | | Db | XX | allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; | | | |
| | | | | Db | XX | bone damage; cartilage damage; antiinflammatory disease; coagulation; | | | |
| | | | | Db | XX | thrombosis; contraceptive; ss. | | | |
| | | | | Db | OS | Homo sapiens. | | | |
| | | | | Db | XX | WO200058473-A2. | | | |
| | | | | Db | XX | | | | |

```
Db 68705 TC---TTTCGATGCGCGGATCAGGACCGGATGCGGGCTGATTTCGACGAACACCCGGTAT 68649
Qy 1474 rGSerLeuIleGlySerProGlyArgThrPheProValHisProLeuAspValMetA 1494
Db 68648 CCGCGTGAAGGCGGTTCGGACGGCGCTTC-----GAATCAGCGGCTGG 68601
Qy 1494 laaPalaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLeuSerArgProGlyT 1514
Db 68600 CGGATGCT----- 68593
Qy 1514 hralaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuG 1534
Db 68592 -----TCGGTACAGTACTCGCGGTTCACACCG 68565
Qy 1534 lYlYsProArgGln-SerProLeuThrTyr-GluAspHisGlyAlaProPheAlaGlyHis 1553
Db 68564 GCGGTATCCATGAGTTCGCCGCTCAGAGTGAG----- 68532
Qy 1554 LeuProArgGlySerProValThrMetArgGluProThrProArg-----LeuGln 1570
Db 68531 -----AAGAACGCCACCGTGAAGTACGGGTTTCGATACCTCGCAGCGCGCGATGAGC 68478
Qy 1571 GluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArg 1590
Db 68477 TCCTCCCGGATCGGTCACCTCGCGGAGTGTGACGCGTAGTCGACGTCGATCTCGCGG 68418
Qy 1591 GluIleAlaLysSerProHis-----SerThrValProGluHisPro-----His 1606
Db 68417 CACGATGCCCTTCGGCTTCATCGTCGTCGATCAGTCCGTCACGGCATCCGTCGCGCG 68358
Qy 1607 ProIleSerProTyrGluHis-----LeuLeuArgGlyValSerGlyValAspLeuTyr 1624
Db 68357 CCAGCAGCAGCAGCAGACACCATTTGACTGACGATATTTCAGTCGGT----- 68310
Qy 1625 ArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp 1644
Db 68309 -----CTCCCATTTGGGACGCCAATTCT 68286
Qy 1645 AlaAlaAlaAlaTyrTyrLeuPro-ArgHisLeuAlaProAsnProThrTyrProHisLe 1664
Db 68285 CGGCTCGGCTGCCACAGGCCAACGAGACCATTCGCGCGGACCGCCCAACCGCACCA 68226
Qy 1664 uTyrProProTyrIleuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnTh 1684
Db 68225 GCAACCG----- 68219
Qy 1684 rIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAl 1704
Db 68218 -----GTCGCGCAGTGCCACCCAGCGCGCGCTCCCGAAGCG 68181
Qy 1704 aGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTy 1724
Db 68180 ACAGGGCGCC-----CGCACATATCCCGCGGAT-----CTCCCTTCGGAAT 68136
Qy 1724 rAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro-----Va 1742
Db 68135 -GCCCGATGACCGGT-----CGGGAACCAACACCCAGCCGACCGCCAC 68095
Qy 1742 lLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPr 1762
Db 68094 AATTCCGCCACGACACCATTCACCGCCCAACCG-----CC 68059
Qy 1762 oThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyPr 1782
Db 68058 GGCTGACCAATCCACC-CGATCC-----AGCCCCGGTGGCC 68021
Qy 1782 oThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAr 1802
Db 68020 G-----GGTCCCGCGCAGCAGCTCGAGCAGCAGCACTCC----- 67984
Qy 1802 gGluArgAspArgGluArgGluLysSerIleLeuThrSerThrThr-----ThrVa 1821
Db 1821 lGluHis-----AlaPro----- 1825
Db 67924 AGCACATAACTGGGCTCCCATGCCAGCACTCGCGCCCTGCGCCCGGGAACAGAAATGC 67865
Qy 1826 -----IleTyr---ArgProGlyThrGluGlnSerSerGlySerGlySerGlyse 1840
Db 67864 CGTCTTGGCCACCGCTGGGCGCGCGCCGACACAGCTCGCGCGCGGACAG----- 67813
Qy 1840 rSerGlyGlyGlyGlySerSerSerSerAlaSerHisSerHis----- 1856
Db 67812 -----CCGCCCCAGACCCGTCAGCCC-TTCCATCAGCTGAGTGC 67773
Qy 1857 -----AlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnAlaAr 1874
Db 67772 GGTGCGGCGCCACGACACCGCCCGATG-----ATCAACACAGCG 67734
Qy 1874 gProSerValLeu-----HisAsnThrGlyMetLysGlyIleIleThrAl 1889
Db 67733 ACCGGGTGTGA-CCAGCGACCAACCCACATCAGAGCGGCGCACGT-----TCCTCA 67684
Qy 1889 aValGluProSerLysProThr---ValLeuArgSerThrSerThrSerProValAr 1908
Db 67683 TCGGCGCCACCCAGGCGCAACCGCTCGCGCTGTGTGTCACCGTTGCGCGACCTC 67624
Qy 1908 gProAlaAla---ThrPheProAlaThrHisCysProLeuGlyGlyThrLeuAspG1 1927
Db 67623 GCCGACAGCACCCACCGCACCGCCCGGTCTGTC-----CATGG 67582
Qy 1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAr 1947
Db 67581 GC CGCGCCAAACCGGTCTCACTGCGGTGCTCTTCCAGAAATCACATGTGCGTGGTCCG 67522
Qy 1947 g----- 1947
Db 67521 CTGATCCGAAACAGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67462
Qy 1948 -----ProGluArgProArgAlaAspThrGlyHisAl 1958
Db 67461 GGCTCGTCAACAACAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67402
Qy 1958 aPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysG1 1978
Db 67401 ACGTGCAACGTCTTGGCGACACCGCGTGGCGCATCGCTGCGCACCATCT-----TG 67351
Qy 1978 ySerGluProArgProLeuValProValSerGlyHisAlaThrIleAlaArgThrPr 1998
Db 67350 ATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67291
Qy 1998 oAlaLysAsnLeu-----AlaProHisHisAlaSerProAspProProAlaPro-- 2014
Db 67290 AGCCACAAAGCGCGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67231
Qy 2015 -----ProAlaSerAlaSerAspProHisArgGlyLysThrG1 2027
Db 67230 ATGGGATCCCCAACGTCGTCGCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCG 67176
Qy 2027 nSerLysProPheSerIleGlnGluLeuArgSerLeuGlyTyrHisGlySerSe 2047
Db 67175 -----CGACACTAACCGCGCACCTG-GCCAG 67151
Qy 2047 rTyrSer-----ProGluGlyValGluProValSerProVa 2059
Db 67150 CGCGCGAGTGATCACCGTTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67097
Qy 2059 lSerSerProSer---LeuThrHisAsp-----LysGlyLeuProLy 2072
Db 67096 GGAGGCGCGCTCTGTGATTGACCGCGCATCCGCGCACCCAGCGCGCGCGCGCGCGCGCA 67037
Qy 2072 sHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro-G 2092
Db 67036 CGCGCGCGCTCAGCCAAACCGCGCCA--GCACGAGACCCCGCGCGCGCGCGCGCGCGCG 66979
```

| | | | | | | |
|----|-------|---|-------|-------|--|-------|
| QY | 820 | oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProValGlu--G1 | 839 | 1125 | leSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProV | 1145 |
| Db | 70321 | ACATCCGCAAGATGCCCCACACACCGGCAACCAACCGCAG-CCCCACCGCAGCCGCA | 70263 | 69465 | -----CCAAACCCCGAAACCGCGCCATACCATACATACCCCGCGCG- | 69424 |
| QY | 839 | uGlyGlu----- | 841 | 1145 | aiGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheS | 1165 |
| Db | 70262 | AGGCAAGTCAACCCGCGCTACCGGGGCAACCGTGGCGCATATACCGCGCTCGAGC | 70203 | 69423 | -----CCAAACACCT--GATACCGCTGGCGACGTCCA--T | 69393 |
| QY | 842 | -----GluGlnLysPro--ProAlaAlaGluGluLeuAlaVa | 853 | 1165 | erGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuG | 1185 |
| Db | 70202 | AACCACCACTGCGGCTCACCGGAGCTGGCGCAACCAACCGCATCTCTCGACGGCAC | 70143 | 69392 | CGCGCGTGGCCCAACCGCGCGCACACCG-ACAGTCCCGCGCGCGCTGCACCGA | 69339 |
| QY | 853 | laspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGlu-- | 872 | 1185 | lyValProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeu-----GlySerV | 1203 |
| Db | 70142 | CGACCATCGGTATCCACCAACACACCGCGCGGATGTTGAGCTGGCGGACAGCGCAC | 70083 | 69338 | GCAACCCCAACACGCGCTGG-CGTGCAAGCTCCATCCCGGACTCGAATC | 69288 |
| QY | 873 | -----GlyProAlaLysGlyLysAlaGluAlaAlaGluAlaThrAl | 887 | 1203 | alProGly-GlySer--IleThrLysGlyIleProSerThrArgVal-ProSerAspSe | 1221 |
| Db | 70082 | CAACCCCAACCGGACGACCGCGCAATCGTGACATCGTCATCAACCGGACCCACCGC | 70023 | 69287 | CGCTGGCGCGGAGCGCAATACACACACATTCGAGTACCCGACTGTCGACCGCGCC | 69228 |
| QY | 887 | agluGlyAlaLeuLysAlaGluLysGluGlyGlySerGlyArgAlaThrAlaLy | 907 | 1221 | ralaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLySg1 | 1241 |
| Db | 70022 | CCCA-----CGGCTCACCAACACACCC | 70002 | 69227 | CACCAACCTGAGCCCGCACCCCATCTCGCGC----- | 69194 |
| QY | 907 | s-----SerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAlaAs | 924 | 1241 | yThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAs | 1261 |
| Db | 70001 | CAGACCGGTGCGATCGTAGCCACACGAGCTGCAACACCTTCAACACCTCGTGACAC | 69942 | 69194 | ----- | 69194 |
| QY | 924 | pGluValaspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro-ArgProSer- | 943 | 1261 | pSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrG1 | 1281 |
| Db | 69941 | CGGTATACACCGTGGCAGC-----ACATCCCGCGCGCTGGCG | 69903 | 69193 | -----AGCACCA | 69186 |
| QY | 944 | -----LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysP | 960 | 1281 | uGlyGlyMetSerValThrGlnCysSerLysGluAsp----- | 1293 |
| Db | 69902 | GCGCGGTGAGCGGCGAGCTCCACA-----CCAGCGCTGCTCGCAATGTCGCGCC | 69849 | 69185 | CGTGCTGCAACGTCACCTCTGACGACCCGAAACACCCACCTCTGACCGCGCGCAA | 69126 |
| QY | 960 | roLeuAspLeuLysGlnLysGlnArgAlaAlaIlePro--ProIleGlnValT | 979 | 1294 | -----GlyArgSerSerSerGlyProProHisGlnThrAl | 1305 |
| Db | 69848 | CT---CCACGCGCAGCGGTGTCGACATCTCTCCAGCAATCCCGGTGCGCGCGGTGCA | 69792 | 69125 | GCCAACTCAACAAACCCGCTGCGCGGCAACACCCACCGCGCGCGCGCGCGCC | 69066 |
| QY | 979 | hrlsValHisGluProArgGluAsp-----AlaAlaProThrLysProA | 995 | 1305 | ala-----ProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAl | 1321 |
| Db | 69791 | CCTCGAACAGCGCGCTGAGACCGGACCACTCACTGCGGACCGGACACACCGG | 69732 | 69065 | AGCCACGCGCGGAGGACCCGATATCCGCGCGCTGCAACACCCACCGCGAGTGGGC | 69006 |
| QY | 995 | laProProAla-----ProProProGlnAsnLeuGlnProGluS | 1009 | 1321 | alleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisS | 1341 |
| Db | 69731 | CAACCCCTGGGATCGGCAACTCCACCGACACCGCGCGCGCGCGCGCGCGCGCTCG | 69672 | 69005 | CGTT-----GCACCACTGACGCAACAATCTCTGCGCGCGCGCGCGCGCGCGCG | 68952 |
| QY | 1009 | erAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProP | 1029 | 1341 | rProHisLysLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProAr | 1361 |
| Db | 69671 | CACACGACCGCGCGAGCCC-----CGCTGCTGCAACACACACACCC | 69630 | 68951 | CCAGATGCG---CGCCACACAGCACCTAGCGCGCTAGCCAGAACCGCTGCGC | 68903 |
| QY | 1029 | roAlaAspLysGluAla-----PheAlaAlaGluAlaGlnLysLeuProGlyAspProP | 1047 | 1361 | gSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArgG1 | 1381 |
| Db | 69629 | CTGCCAGGAAACGCAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 69580 | 68902 | -----CGCAA-----ACGCGTACGTC | 68886 |
| QY | 1047 | roCystTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerP | 1067 | 1381 | uGlyThrProProProProProProProProProProProProProProProProPro | 1401 |
| Db | 69579 | -----ACGCATGCAACCGCGCATCCACACCGCGCGGATGAA----- | 69544 | 68885 | GGCACTCCACCGCGCGCTCCCGAGTGGCAAAACCGCGCGCGCGCGCGCGCGCGCG | 68826 |
| QY | 1067 | roHisAlaProAspProSerAlaPheSerTyrAlaProGlyHisProLeuProLeuG | 1087 | 1401 | la-----LeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaT | 1417 |
| Db | 69543 | -----TCCCAACCCCGGTAT----- | 69528 | 68825 | CCCGACAGTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 68766 |
| QY | 1087 | lyLeuHisAspThrAla-----ArgProValLeuProArgProProThrIleSerAspN | 1105 | 1417 | hrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrP | 1437 |
| Db | 69527 | CGGCACACCTCTAGGAGAGTACGTCGCGGAAACACCTCGCGCGCGCGCGCGCGCGCG | 69468 | 68765 | CCAGCGT-----CGGAATGCAATCGGTTCAC-- | 68737 |
| QY | 1105 | roProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaI | 1125 | 1437 | roGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProL | 1457 |
| Db | 69467 | CT----- | 69466 | 68736 | -----CGTAGCGCGCGGTACACGTGACGACCAACGTC | 68706 |
| | | | | 1457 | euLysTyrAspThrGlyAlaSerThr-----ThrGlySerLysLysHisAspValA | 1474 |

[illegible]

are useful in diagnostics as expressed sequence tags (EST) for
identifying expressed genes or for physical mapping of the human genome,
in forensics, in assessing biodiversity, or in identifying mutations
responsible for genetic disorders and other traits. The nucleotide
sequences are also useful as hybridisation probes, as oligomers for PCR,
for chromosome and gene mapping, in the recombinant production of
protein, or in generating antisense DNA or RNA. The purified polypeptide
is useful for generating antibodies specific for it. The present sequence
is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
for this patent did not form part of the printed specification, but was
obtained in electronic format directly from USPTO at
cc seqdata.uspto.gov/sequence.html?DocID=20030073623

XX Sequence 499 BP; 183 A; 80 C; 123 G; 111 T; 0 U; 2 Other;

Alignment Scores: 1.71e-10 Length: 499
Pred. No.: 585.50 Matches: 117
Score: 85.06% Conservative: 14
Percent Similarity: 75.97% Mismatches: 22
Best Local Similarity: 4.43% Indels: 1
Query Match: 8 Gaps: 1
DB:

US-09-522-753-5 (1-2517) x ACH32484 (1-499)

QY 270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289
DB 38 ATATTGACTAACAGGTGATGAGCAAACTATTTTATTTTAAAGAAATCAT 97
QY 290 AlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309
DB 98 GCAGAAACAAAGGAAACAAATACTCTGAGGTTATGATCAGCTCATGAGGCGATGG 157
QY 310 GluLysLysValGluArgIleGluAsnProArgArgAlaLysGluSerLysVal 329
DB 158 GAGAAAGTGGACAGATANAATAATCTCTGTGAAAGCTAAAGAAAGCAACA 217
QY 330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg 349
DB 218 AGGGAATCTATGAAAGCAGCTTCCAGAAATTCGAAACAAAGAAACACACAGNAGA 277
QY 350 MetGlnSerArgValGluArgGlySerGlyLeuSerMetSerAlaAlaArgSerGlu 369
DB 278 TTTCAG--CGATGGGCATAGGGAGCTGCTTTTCAGCCACCATTGCTAGGAGTGAG 334
QY 370 HisGluValSerGluIleLeuAspGlyLeuSerGluGlnGlnGluAsnLeuGluLysGlnMet 389
DB 335 CATGAGATTCTCAATATTGATGGCTCTCTGAGCAGGAGAAATAATGAGAAACAAATG 394
QY 390 ArgGlnLeuAlaValIleProMetLeuTyrAspAlaAspGlnGlnArgIleLysPhe 409
DB 395 CGCGCGCTCTCTGTGATTCACCTATGATGTTGATGACAGAAACAGAGCCAGTTC 454
QY 410 IleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyr 423
DB 455 ATTAACATGATGGCTTATGAGGAGCCCTATGACAGTGTAT 496

RESULT 97

AAI99683_18/c
Continuation (19 of 44) of AAI99683 from base 1800001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

| Fragment Name | Begin | End |
|----------------|---------|---------|
| WP AAI99683_00 | 1 | 110000 |
| WP AAI99683_01 | 100001 | 210000 |
| WP AAI99683_02 | 200001 | 310000 |
| WP AAI99683_03 | 300001 | 410000 |
| WP AAI99683_04 | 400001 | 510000 |
| WP AAI99683_05 | 500001 | 610000 |
| WP AAI99683_06 | 600001 | 710000 |
| WP AAI99683_07 | 700001 | 810000 |
| WP AAI99683_08 | 800001 | 910000 |
| WP AAI99683_09 | 900001 | 1010000 |
| WP AAI99683_10 | 1000001 | 1110000 |

WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 140000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Alignment Scores: 1.16e-08 Length: 110000
Pred. No.: 584.50 Matches: 647
Score: 30.84% Conservative: 255
Percent Similarity: 22.12% Mismatches: 944
Best Local Similarity: 4.42% Indels: 1092
Query Match: 4 Gaps: 148
DB:

US-09-522-753-5 (1-2517) x AAI99683_18 (1-110000)

QY 58 IleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsn--- 76
DB 72287 GTAGTCGAAGATCAGCGTGGCGA-----CAACGTGAGCCCGTGGCGGT 72243
QY 77 ---GluArgSerGlnGluLeuHis-----LeuArgProGlu----- 87
DB 72242 CTTGAGTCGGTGGCGAGCTCCAGCGCGGTCAAGGAATCGAACCCGAGATCTTGAATGC 72183
QY 88 -----Ser 88
DB 72182 TTTGTGAGCGTTGATATCGACACTGTGGAACGACCCCAACCCCGCGCATGTGTGAT 72123
QY 89 HisSerTyr-----LeuProGluLeuGlyLysSerGluMetGluPheIleGlu 104
DB 72122 CACCAAAATCGGTAGTTGCGCGCTCCCGCGCTCGGCGGTAGCGCATGCGAG----- 72072
QY 105 SerLysArgProArgLeuGluLeuProAspProLeuLeuArgProSerProLeuLeu 124
DB 72071 -----CGGCTGA-CTAAGC 72058
QY 125 AlaThrGlyGlnPro-----AlaGlySerGluAspLeuThr 136
DB 72057 CCGACATCGACACCGCGGTATCGCGCGGTGCGATCTCCCGCGGTGCG----- 72010
QY 137 LysAspArgSerLeuThrGly----- 143
DB 72009 ---GACCCGCGAGCCACTGGTGTGTAATATCGGTAGCGAGCAGCAGCGCGCGCAGCG 71953
QY 144 LysLeuGluProValSerProSerProProHisThrAspProGluLeuGlu----- 161

| | | | |
|----|------|---|-----------------|
| Db | 7708 | ATCACAGCGCAGAG----- | ATCACAGCGG 7731 |
| Qy | 1561 | ThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSer-----Ser 1577 | |
| Db | 7732 | GCTCAGTCTACTCCATCTCCAGCTCTCCGCCAGACACAAAGGCTCTGATGTTCACACC 7791 | |
| Qy | 1578 | LysAlaSerGlnAspArgIysLeuThrSerThrProArgGluLeuAlaLysSer-ProHi 1597 | |
| Db | 7792 | AGCTCCAGCACCCCTGAGGAAGATTCTCATGAGCCCAAGTATGTG-----TCTGCCACA 7845 | |
| Qy | 1597 | sSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgG1 1617 | |
| Db | 7846 | AGTGTCACCTCCACAAGTGTCCACACAGCATTCGACAGCCTG-----7888 | |
| Qy | 1617 | yValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerI1 1637 | |
| Db | 7889 | -----TCAGTGTCTGCCCTTGCTCTAC---ATGAGGCCCG 7920 | |
| Qy | 1637 | eProArgGlyIleProLeuAspAla-----Al 1646 | |
| Db | 7921 | CCCCGCCAGTTCACCTCTAAAGACCTTTAGAGAAAAAACAGCACCTCCAGTGCACAAAC 7980 | |
| Qy | 1646 | aAlaAlaTyrLeuProArgHis-----LeuAlaProAsnProThrTyrPr 1662 | |
| Db | 7981 | AACCTCGAGATACAGCCTCGAGGTGCTGGTAGCTGTGCACAAAGAAAGGTGGCTCCA 8040 | |
| Qy | 1662 | oHisLeuTyrPro-----ProTyrLeuLeuArgGlyTyrProAspThrAlaAlaLeuG1 1680 | |
| Db | 8041 | GTCAATTGCTCCCAAAATTACCTCTGTATTAGCCGGATG-CCTGTGACGATTGACCTGGA 8099 | |
| Qy | 1680 | uAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAl 1700 | |
| Db | 8100 | AAATTCCACGAAGATA-----8115 | |
| Qy | 1700 | aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLe 1720 | |
| Db | 8116 | -ACCTTGGCAAAACCAGCTCCTCAAAACCTCCTCACTGGTCTG-----8154 | |
| Qy | 1720 | uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLe 1740 | |
| Db | 8155 | -----GTGAGCGCACTCACTGCGCTGGTGAACGTCTCCGTGCTCCG---GT 8198 | |
| Qy | 1740 | uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760 | |
| Db | 8199 | GAATGCCCTGAAAGGCCCGCTGAAGGCTCAGTGACCACTGAAAGTTGTGTGAGC---8256 | |
| Qy | 1760 | rLeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSerProGl 1780 | |
| Db | 8257 | -----ACCCCTGC 8264 | |
| Qy | 1780 | yGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspAr 1800 | |
| Db | 8265 | TGGGCCCGTGAAACGCTCTCGAAAGGCCCTGTG-----8295 | |
| Qy | 1800 | gAspArgGluArgAspAspArgGluArgGluLysSerIleLeuThrSerThrThrTh 1820 | |
| Db | 8296 | -----AATGTTCTTACGGGCCAGTGAA 8318 | |
| Qy | 1820 | rValGluHisAlaProIle-----TrpArgProGlyThrG1 1832 | |
| Db | 8319 | TGTTCTCACCCTCCAGTGAACGCCCGTGGGCACAGTGAATGCCGCCCCAGGCACAGT 8378 | |
| Qy | 1832 | uGlnSerSerGlySer-----1837 | |
| Db | 8379 | CAATGCCGCTCGAGTGCAGTGAATGCCACAGCAAGTGCAGTACCGTCAACAGCGGTGC 8438 | |
| Qy | 1838 | ---SerGlySerSerGlyGlyGlyGlySer-----1847 | |
| Db | 8439 | GTTACTGCTGCATCTGTTGTGTAAACGCCCAACAGGCACGCGTGCATGCGACATGGCAGGGC 8498 | |
| Qy | 1848 | ---SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProI1 1863 | |

8499 AGTGATTGCCCGCTCAACAAAGTGCACAAAGAGAGCGAGTGTCTAATGAAAAACAGTCGGTT 8558

1863 eSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHisAsnThrGlyMet 1883

8559 CCACCCAGGGTCCATGCTGTGATGACGATCGTCCGCCA-----GACGCGGCGC--- 8607

1883 tLysGlyIleIleThrAlaValGlnProSerLysProThrValLeuArgSerThrSer-- 1902

8608 -TCAGGCGCGGGCGTGGTGTGAACACTTCTGAAGGGGTGTGCTCTCTGAGTTACTCAGG 8666

1903 -----ThrSerSerProValArgProAlaAla-----ThrPheProProAlaTh 1917

8667 GCAGAACGACGAGGCCCAAGCGGATCAGCGCCCAAGATCAGCCAGATCCCCCGGCGCAG 8726

1917 rHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLe 1937

8727 T-----GCAATGGACATTAATTTTCAGCAGTCTGAGTGTCCAAAGTCCCA 8768

1937 uLeuProLysGluAlaProArgValAlaArgPro-----GluArgProArgAlaAspTh 1955

8769 GGTCAAACTGATTTCTGTACAGATCGAGCTTCATCCAAAGGCCCTCAAGTCTCTGC 8828

1955 rGlyHisAlaPheLeuAla-----Ly 1962

8829 AGGCTATGCGAAGCTGCCACCCATTCCACGTTGGTACTGACGCCCGCAGACATATAATGC 8888

1962 sProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProAr 1982

8889 CTCCTCTGTGATTTCTGTGTGAAG---GCCGATAGGCCATCTCTTG---GAGAAGGCCGA 8942

1982 gProLeu----- 1984

8943 GCCCATTCACCTCTCGGTGTCCACGCGCTGTACCCAGGGAGGCACAGTGAAGTTCTCAC 9002

1985 -----ValProProValSerGlyHisAlaThrIleAlaArgThrProAla-- 1999

9003 CCAGGGGATCAACACACCCCTGTGTGGTTCACACACAGCTGTCTCTACCCCAAGCAT 9062

1999 ----- 1999

9063 TGTACACAAACAAAGAGCTTGCTGACCCCGCTCACCTTTAAAATCGAGACCAAGGTCTC 9122

2000 -----LysAsnLeuAlaProHisAlaSerProAspProProAl 2013

9123 TCAGCGCGCCCACTGGGGTCCACGCTCCAGCCCGCCACAC-----CCTCTCTGC 9170

2013 aProProAlaSerAlaSerAspPro-----HisArgGluLysThrGlnSerLysProPh 2031

9171 TCTGCCAGCAAACTGCCCTACAGAGTCAACCATGTCCCTCGGGGCCAGCATCCCGCAGC 9230

2031 e-----SerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSe 2049

9231 AGATCGAACTGTCTCCATTTTGGCAGCTGCAAGAGTAGATGCTCAT-----TC 9278

2049 rPro-GluGlyValGluProValSerProValSer-----SerProSerLeu 2065

9279 TCTTCGCAAGTGCACCCGCGCCATCTCATTTCCCAAGGCAAGCCACCCAGCAGTAC 9338

2065 hHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyG 2085

9339 TGCAT----- 9343

2085 lLeuArgProLysGlnProGlyProValLysLeuGlyGlyAlaAlaHisLeuProH 2105

9344 --CTACGGCGCTCTCCACCAACGCCACAGTATGTGGCTCAGAGCATCCCAAGTGCACCA 9401

2105 iLeuArgPro-----LeuProGluSerGlnProSerSerSerProL 2119

9402 GTTCATCTCCAGCATCCACCCAGACAGTCTGTCTCATCATGCCACCCACACATCACCCA 9461

2119 euLeu-----G 2121

9462 GACTGTGTCTCCAGCACCCTCTCCAGCGCGAGGTGAGATGACATCTCCACAGCTGCC 9521

| | | | | | | | |
|----|------|--|------|----|------|--|------|
| Db | 3766 | AAAGTCTGAGAGTTTGGCAGTCTCTAAAGAAAGATGTAGATGAATATGAAGAGCGTAGCCTC | 3825 | Db | 4840 | CTAGAGAGAAAGAGGAGAGATTCTGACTTCTGCTAGGATCTATGGAGAGCAGACA | 4899 |
| Qy | 526 | GluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeu | 545 | Qy | 685 | LysAlaProAlaAlaLysSerGluGluAlaAlaPheProProValVal | 700 |
| Db | 3826 | GTTTCAGAGGTAGGCAAAACCCCTCAA | 3879 | Db | 4900 | TCTGAGGAGCAAAACAGCACAACTGATTCCTCAATCAAGAACCCAGTAGTCTCTTCCATAGC | 4959 |
| Qy | 546 | LeuLysGluLysThrAspAsp | 558 | Qy | 701 | -----GluAspGluGluMetGluAlaSerGlyVal | 710 |
| Db | 3880 | AAAAAGAAAGGATGGATCATGTGATTTGATATCTGCACCAAGCGAGAACGGAATTC | 3939 | Db | 4960 | AGATTATTTGAGAGTCTACACGGATGCAACAGAAAGAAAGAAAGAACCCAGAACCCAAA | 5019 |
| Qy | 559 | AspGluLysGluAlaValAlaLysLysGlyArgLysThrAlaAsnSerGlnGlyArg | 578 | Qy | 711 | SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu | 730 |
| Db | 3940 | AGAAATTCACGCCCAATCAGCGAAGATTCTGAAAGGACTGGTGTCTCCACGATGTCGGA | 3999 | Db | 5020 | GAGTTTGAAACAGCAAGATACAGAGATATCATCCAAAGACCCCAAGATCTGCTCTCAG | 5079 |
| Qy | 579 | LysGlyArgIleThrArg | 587 | Qy | 731 | ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThr | 748 |
| Db | 4000 | CATGGTTCCTTCATGAAGATGAGATCCCATAGGTCCTCCCTAGGCTACTGTCACTAAA | 4059 | Db | 5080 | AATAAAGATTTCAGAACTGAAAACTCCACTTCCTGTTGGCCCTCCAAGTGTACAGTCGTA | 5139 |
| Qy | 588 | AsnGluAlaAsnSerGluGluAlaIleThrPro | 598 | Qy | 749 | -----GluSerIleProSer | 753 |
| Db | 4060 | GGGTCTCTAAAGTAGATGAAAGTCTCCCTATTCTAACATACAGTCAGGGAAGAG | 4119 | Db | 5140 | ACTCTAGATACAGCCCATCAGCACTAGAGAAAGACCACTGGTGACAAAACGGTAGAGCG | 5199 |
| Qy | 599 | -----GlnGlnSerAlaGluLeuAlaSer | 606 | Qy | 754 | -----ProHisThr | 761 |
| Db | 4120 | TCTTTAAAAATTAATCCCTTATGATCTTAGCAGGAGAGACAGATGGCAGATATGGCCAAA | 4179 | Db | 5200 | CCTTTGGTAAACAGAGAAGACTGTGGAGCCAGCTACCGTCTCAGAGAACCAAGCCT | 5259 |
| Qy | 607 | MetGluLeu | 620 | Qy | 762 | ThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProGly | 781 |
| Db | 4180 | ATPAAACTATCTGTCTTGAATCTGAAGATGAATCTGGAATCTGAGTCTCAGATGAAA | 4239 | Db | 5260 | GCATCTGAACCTGCTCTCCCTCTGTGGAAACAGCTGGAACAAGTAGACCTGCCCCAGGA | 5319 |
| Qy | 621 | GluThrAla | 623 | Qy | 781 | ----- | 781 |
| Db | 4240 | CAGGATGCTGGCAGATTTGATGTGAGTTTCCCAACAGCATATTAAGAGAGATAGCCTT | 4299 | Db | 5320 | GCAGACCCCGATAAAGAGCTGCCATGATGCTCGGGTGTTCAGGAAAGTTTCATCAGT | 5379 |
| Qy | 623 | ----- | 623 | Qy | 782 | -----ProThrThrProProArgArgThrSerArgAla | 792 |
| Db | 4300 | CGAAAAAGGTGTGTACGAGATCTGGAACCTGGTGAAGTCTCTGATTTCTGACGAAGAT | 4359 | Db | 5380 | GACCAACCCGCTTATCTGATGATCCAAAGCTCTCAACTCCCGGGCCTCGTTTCCAGGCA | 5439 |
| Qy | 623 | ----- | 623 | Qy | 793 | Pro-----IleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro | 810 |
| Db | 4360 | GGTGAACACAAATCCCACTCACCCAGAGCCTCTGCATATATGAAAGTTCTCGATGTCT | 4419 | Db | 5440 | GAGAGCAACGTAGATCCAGAGCCTGACAGT | 5481 |
| Qy | 623 | ----- | 623 | Qy | 811 | AlaProProProSerAlaProProProValValProLysGluLysGluGluGlu | 830 |
| Db | 4420 | TTTTTATTGAGGACAGAGAGACAAGCTACGTGAGGAGATGAAGAACTCTTAGTTCT | 4479 | Db | 5482 | TCAAAACCACTCAGAGTCTGAGAGCAACCAATGAGCAAGCCGCAAGAACCCAGAGCC | 5541 |
| Qy | 624 | -----LysLys | 625 | Qy | 831 | ThrAlaAlaAlaProPro | 845 |
| Db | 4480 | TTAGAAAGGAACAAATTTTACTCTTTTGCATTGGATAGACAATCACACCAAGACTAAA | 4539 | Db | 5542 | ACTGAGATGCTGAGCTGATGCAACCAAGCAAGCTGCTCTCTGAGTCTCAGCCC | 5601 |
| Qy | 626 | GlyLeuLeuGluHisGlyArg | 637 | Qy | 846 | ProAlaAlaGluLeuAlaValAsp | 862 |
| Db | 4540 | GCTTTGCTTGAAGAGCTAAATCCCTCTCTCATCTCGTGAAGAAATTTGTCCTTCTT | 4599 | Db | 5602 | CCAGCTTCTCAAGATTTAGAGTTGATCTCTCCAGTTGCTGCAAGGATATAAAGCCAAAC | 5661 |
| Qy | 638 | -----AlaArgMetValGlySerLysThrValSerGln | 648 | Qy | 863 | LysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAla | 882 |
| Db | 4600 | GATTTGGACTCCCGATTGCAATTTTCGAACACAAAGATTAAGAAAGGTTGACTCT | 4659 | Db | 5662 | AAAAAGCAAGCGTTCAAGACCCCTGTTGAGCAGCTGATGATGATGATGATGATGATGAT | 5721 |
| Qy | 649 | -----CysLysAsnPheThrPheAsnThrLysLys | 659 | Qy | 883 | AlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArg | 902 |
| Db | 4660 | GCTCCAGACCTATTCATCTCTGATCATGAAAGAAAGAAAGAAATTTAGGACTGATTCAGAA | 4719 | Db | 5722 | GTCAACAGGAAGAGTGAAGGATAGACCGGAAAGAACTCAAGCGGTCTCAATTTCTCTCGG | 5781 |
| Qy | 660 | GlnAsnLeuAspGluLeuGlnGlnHisLysLeuLysMetGluLys | 675 | Qy | 903 | AlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSer-----SerAlaThrCys | 921 |
| Db | 4720 | GGGAAATGATGATATAAGAAAGAGGACCATATAAAGAAAGAGAGAGAGAGGAGGAGGATG | 4779 | Db | 5782 | CGAAGACACAGAAAGCTTTTGGAAATTTGAAGATGAGGAGGAGAGAGATACAGGACTGCT | 5841 |
| Qy | 676 | -----GluArgAsnAlaArg | 681 | Qy | 922 | SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg | 941 |
| Db | 4780 | TTTGCTTCTGTTTTTACAGCTCAATCTTTGAACAAAGATTTCCAAAGCGATTCAGCAT | 4839 | Db | 5842 | TCTAAAAACTCTGCTGAGACCTTGAA | 5877 |
| Qy | 682 | -----LysLysLys | 684 | Qy | 942 | ProSerLeu-----LeuThrProThr | 948 |
| Db | 4839 | ----- | 4839 | Db | 5878 | CCNAGTTTGGCTCTCAGCCGAAACCAAGCGCGGAAATGTAAAGAGCGTCTATGCAACCATG | 5937 |

CC is afflicted with breast cancer. The method comprises comparing the level
 CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
 CC ABR47386 to ABR47632) in a patient sample and the normal level of
 CC expression of the marker in a control non-breast cancer sample, where a
 CC significant increase in the level of expression of the marker in the
 CC patient sample and the normal level is an indication that the patient is
 CC afflicted with breast cancer. The breast cancer associated sequences from
 CC the present invention have cytostatic activities and can be used in gene
 CC therapy. The method is useful for diagnosing and treating breast cancer.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 12227 BP; 3448 A; 3248 C; 3069 G; 2462 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.92e-09 Length: 12227
 Score: 586.00 Matches: 638
 Percent Similarity: 30.91% Conservative: 379
 Best Local Similarity: 19.39% Mismatches: 1071
 Query Match: 4.43% Indels: 1207
 DB: 7 Gaps: 150

US-09-522-753-5 (1-2517) x ACC50291 (1-12227)

Qy 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
 Db 2077 GACTATAACCAAGATCGTACATATTATGAGAGTGTTCGAACCTCCAGGCACCTTATCTCTGAG 2136
 Qy 61 ProGlnArgArg---ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer 79
 Db 2137 GATTCCAGCGGGGACTATCCAGCTCGAGGAGAGAGTTTTATTTCAGAAATCGGAACCTTAC 2196
 Qy 80 Gln----- 80
 Db 2197 CAAGGAGACTACTATGTAATCAGTACTACGATGATCTCGGGAATACAGGGATTACAGG 2256
 Qy 81 -----GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96
 Db 2257 AATGATCTTTATGAACAAGATATTAGG-----GAATATAGTTACAGGCAAGGAACGA 2310
 Qy 97 LysSerGluMetGluPheIleGluSerLysArgProArg-----Leu 110
 Db 2311 GAAAGAGACGTAAGAGATTGAGTCTGACCGGACAGACCATGATGAGAGGCGCGATT 2370
 Qy 111 GluLeuLeuProAspPro-----LeuLeuArgPro-----SerProLeu 123
 Db 2371 GAACGAAGTCAAAGTCTGTCTTCACTTCGCGAGTCCACAGAGTCTCTGGAGCGTCTCCCTCT 2430
 Qy 124 LeuAlaThrGlyGlnProAlaGlySerGlu-----AspLeuThrLysAspArg 139
 Db 2431 CAGGCAGAGAGGTTCGCGAGTATCTCTGAGGAGGCGTTTACAGCGCATCTCTCAGACCGG 2490
 Qy 140 SerLeuThrGlyLysLeuGluProValSerProPro----- 151
 Db 2491 AGT-----GGAAGCTGTAGCTCACTCTCCCTCCCAAGATATGAGAACTGGCAAGTCT 2544
 Qy 152 -----SerProProHisThrAspProGluLeuGluLeuValProPro 165
 Db 2545 CGTTTGGAGCGCTATACAAAAATGAAAGACAGATAAAGCAAGCACTTTTGATCCGGAG 2604
 Qy 166 ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMet 185
 Db 2605 AGAGTGGAGAGAGAGACGCTTAATACGGAAGAAAAAGTGGAAAAAGGACAAACTGAC 2664
 Qy 186 ValGluGlnGlnIleSerLysLeu-----LysLysLysGlnGlnGlnLeuGluGlu 202
 Db 2665 AACGACAAACGCAAGGAAAGGTTTCACTCCCTCTAGTTCTTCCAGTCTTCCAGAAACGGACCA 2724
 Qy 203 GluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu----- 220
 Db 2725 GAAATGAGGAGAGCAAGCCCTGAAAGCCCGAGGAGTTGTAATAACTGACGAGAGAG 2784

Qy 221 -----SerLysHisArg---SerLeuValGlnIleIleTyrAsp 232
 Db 2785 AAAGCTGCAAAAGAGGGAATAGCGAAAAACCGCTCGAACTCATCCCTCGTGGTGTG 2844
 Qy 233 GluAsnArgLysLysAlaGluAlaAa-----HisArgIleLeuGluGlyLeuGly 249
 Db 2845 ACTCGAGTGAAGAGAGAAAGAGAAAGGTCATTGACCACTCTCTGTGGAAAGATTGAAA 2904
 Qy 250 ProGlnValGlu-----LeuProLeuTyrAsnGlnProSer 261
 Db 2905 GCCAAGCTTGATATGACACTGTCAAAATCTTCTGCCCTGGACCAAACTTCAGGTCTCT 2964
 Qy 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu--- 280
 Db 2965 CAGACGAGCGCTGCAAAATCTGACTTGTCTAAACTGGAATCAGTCTAGAAATGAAATGACCA 3024
 Qy 281 -----IleLeuTyrPheLysArgArgAsnHisAla 290
 Db 3025 AAGGAAAAAGGGGCTTTCAAGCCCATGTTGAAGTGTGGAGAAAGGAGCGCTTAAAGCC 3084
 Qy 291 ArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGlnLeu-----Met 306
 Db 3085 AGGAAGCACTCAAGCCTGAG-----CAGCTGCAGATGGGTAACTGTGTGGATCTG 3138
 Qy 307 GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu 326
 Db 3139 GAGAAGCTGGAAGCCAGG-----AAAAAGCGCTTTGCAGAT 3174
 Qy 327 SerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeu 346
 Db 3175 TCCAAATTTAAAGCA-----GAAAGCAAAACACAGAGTCAAGAAA--- 3216
 Qy 347 GlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAla 366
 Db 3217 -----AGCAGTCCAGATGGAGGATGCT 3240
 Qy 367 Arg-----SerGluHisGluValSerGluIleIleAsp 377
 Db 3241 CGCGTGTCTTCAAAAAAGCAGCCTGACGTGTCTCTAGAGAGGTCTTCTGCTGAGGAA 3300
 Qy 378 GlyLeuSerGluGlnGluAsnLeuGluLysGlnMet----- 389
 Db 3301 GGAGAGCTGAAAGAAAGCCTGTGAGGAAGAAATCTTAAAGAGAGAACTTAAAAAATC 3360
 Qy 390 -----ArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnArg 406
 Db 3361 AAAGTGGACAGACTTAATCTGTGTCAGCCGCCC-----AAAGACTGTCAGAG 3408
 Qy 407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys----- 424
 Db 3409 CTTGCCAGTATTTCTCTTG--GGTCTGGCTCAAGGCCACGCTCAGACCTACAAAGCAAGAC 3466
 Qy 425 ---AspArgGlnValMetAsnMetTyr-----SerGluGlnGluLysGluThrPhe 440
 Db 3467 TGGGAGAACTAGCAGGTGAATCTGTGGAATAATCAAGAGTCCCAATCAAAAAGCCCAATC 3526
 Qy 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
 Db 3527 CCTCAAAACCAACAGCTCAACAGCTCGAGGTATTAGTATGATCAAGGACGAGAGAGAG 3586
 Qy 461 -----LysThrValAlaGluCysValLeuTyrTyr----- 471
 Db 3587 ACCTTAGGAAAAAACTATTGTCAGTCTTCGTGATGAAACACCTGAACCGTAAATCAGGCCAAG 3646
 Qy 472 -----LeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArg 485
 Db 3647 AGAAATCATTACAGTAATACTAGAAAGAAAAAATGGCATTTGATCATCGATCACACGAGA 3706
 Qy 486 SerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 505
 Db 3707 GT-TACCGAAACAAATGGAACAGAGTCTGTAGGAAACAGCAGATGGAATAGGAAATAGCC 3765
 Qy 506 GlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLys 525

| | | | |
|------|----|--|------|
| 1848 | Qy | -----SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIl | 1863 |
| 7417 | Db | AGTGATTGCGCGCTCAACAAAGTCAAAACAGAGAGCGAGTGTCTAAATGAAACACAGTTCGGTT | 7476 |
| 1863 | Qy | eSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMe | 1883 |
| 7477 | Db | CCACCCAGGGTCCATGCTGTGATCGACGATCGTCCGGCA-----GACGCGGGC-- | 7525 |
| 1883 | Qy | tlYsGlyIleThrAlaValGluProSerLysProThrValLeuArgSerThrSer-- | 1902 |
| 7526 | Db | -TCAGGGCGGGCTCGGTGTGAACATCTCTGAAGGGTTGTGCTCTCTGAGTTACTACAG | 7584 |
| 1903 | Qy | -----ThrSerSerProValArgProAlaAla-----ThrPheProProAlaTh | 1917 |
| 7585 | Db | GCAGAAGACCGAAGGCCCAACAGGGATCAGCGGCAAGATCAGCCAGATCCCGCGGCAG | 7644 |
| 1917 | Qy | rHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLe | 1937 |
| 7645 | Db | T-----GCATGGACATTTGAATTTTCAGCAGTCAAGTGTCCAAAGTCCCA | 7686 |
| 1937 | Qy | uLeuProLysGluAlaProArgValAlaArgPro-----GluArgProArgAlaAspTh | 1955 |
| 7687 | Db | GGTCAACCTGATTCTGTACACGATCGCAGCTCCATCCAAAGGCCCTCAAGTCTCTGC | 7746 |
| 1955 | Qy | rglyHisAlaPheLeuAla-----Ly | 1962 |
| 7747 | Db | AGGCTATGCGAAGCTGGGCCACCCATTCCACGTTGGTACTGACCGCCACAGACATATAATGC | 7806 |
| 1962 | Qy | sProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProAr | 1982 |
| 7807 | Db | CTCTCTGTGATTCTGTCTGTGAAG--GCCGATAGGCCCATCTCTTG--GAGAAGCCCGA | 7860 |
| 1982 | Qy | gProLeu----- | 1984 |
| 7861 | Db | GCCCATTCACCTCTCGGTGTCCAGCGCTGTACCCAGGAGGCACAGTGAAGTTCCTCAC | 7920 |
| 1985 | Qy | -----ValProProValSerGlyHisAlaThrIleAlaArgThrProAla-- | 1999 |
| 7921 | Db | CCAGGGGATCAACACACCCCTGTGTCTGTGTTCAACACAGCTGTCTCTCACCCCAAGCAT | 7980 |
| 1999 | Qy | ----- | 1999 |
| 7981 | Db | TGTCAACCAACAAAGAAGCTTGCTGACCCCGCTCACCCCTTTAAATCGAGACCAAGGTCTCT | 8040 |
| 2000 | Qy | -----LysAsnLeuAlaProHisHisAlaSerProAspProAl | 2013 |
| 8041 | Db | TCAGCGCGCCAACTGGGGTCCACGCTCAGCCGCCACCCAC--CCTCTCTGC | 8088 |
| 2013 | Qy | aProProAlaSerAlaSerAspPro-----HisArgGluLysThrGlnSerLysProPh | 2031 |
| 8089 | Db | TCTGCCCGCAAACTGCCTACAGAGTCAACCATGTCCCTCGGGGCCACGATCCACG | 8148 |
| 2031 | Qy | e-----SerIleGlnLeuLeuArgSerLeuGlyTyrHisGlySerSerTyrSe | 2049 |
| 8149 | Db | AGATCGAACTGTCTCCATTGGCAGCTGCMAAGCTAGATGCTCAT-----TC | 8196 |
| 2049 | Qy | rPro-GluGlyValGluProValSerProValSer-----SerProSerLeuT | 2065 |
| 8197 | Db | TCCTCGAACCAAGTGGACCCGGCCATCTCTATTTCCCAAGGGCAAGCACCACCCAGCATAC | 8256 |
| 2065 | Qy | hrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyG | 2085 |
| 8257 | Db | TGCAT----- | 8261 |
| 2085 | Qy | luLeuArgProLysGlnProGlyProValLysLeuGlyGlyAlaAlaHisLeuProH | 2105 |
| 8262 | Db | --CTACGGCGCTCTCCACCAACGCCACAGTCATGTGCTGTGAGGCATCCGAGTGC | 8319 |
| 2105 | Qy | isLeuArgPro-----LeuProGluSerGlnProSerSerSerProL | 2119 |
| 8320 | Db | GTTTCATCTCCGATCCACCCAGAGAGTCTGTATCATGCCACCCCAAGCATCACCCA | 8379 |
| 2119 | Qy | euleu-----G | 2121 |

| | | | | | |
|----|------|--|------|--|------|
| Db | 4796 | CCAAAGTTTGGCTCTCAGCCGGAACAGCGCGCGGAATGTAAGGAGCGCTCTATGCAACCATG | 5813 | GAAAACTTCCAGCAGCTCCACCTTATCTCTGAGAAATCCAGACAGATCTCGAACCCCCC | 5872 |
| Qy | 949 | GlyAspProArgAlaSerProGlnLys---ProLeuAsp--- | 1215 | ThrArgVal-----ProSerAspSerAlaIleThrTyrArgGlySerIleThr | 1230 |
| Db | 4856 | GGTGAC---CATGAAACCCGCTCTCTGTTCAAGAGCCGTTGAGCAACCAAGAGTGACC | 5873 | GCAGGTGCACAGCGCTCGACGCTTCTGAG----- | 5902 |
| Qy | 963 | -----LeuLysGlnLeuLysGlnArgAlaAlaIleProProIleGlnValThr | 1231 | HisGlyThrProAlaAspValLeuTyrLysGlyThrIle---ThrArgIleIleGlyGlu | 1249 |
| Db | 4913 | AGAAAGAGATTGGAGCGAGGCTTCAGAGGCTGCAGCGGTGCC-----ACCACCCCT | 5903 | GAAGGAATGGACAGATGAGGCTGCTATCTGCATCTCGTGAACCTGAGCGCTCTACAGAA | 5962 |
| Qy | 980 | LysValHisGluProProArg-----GluAspAlaAlaPro | 1250 | AspSer-----ProSerArgLeuAspArg | 1257 |
| Db | 4967 | CGAGGGGAAAGCCCTCCAAAGACACAGCGCGGCGAGCCGATGAAGAGGAGGAACGAGGCC | 5963 | TCCTTAGGCTCCAGTCAATGCTCTCCACCTCAGCGCGCCCAACACAGATACCAAGAA | 6022 |
| Qy | 992 | ThrLysProAlaProProPro----- | 1258 | GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisVal | 1277 |
| Db | 5027 | AAGGAACCTGCAGAAACACTCAACCTCAGCGCGGCGAGCGATGAAGAGGAGGAACGAGGCC | 6023 | GCAGAGGAAATAGCAGTGAACCTCACACTCAGTGCAGAGCCCAAGGGGTCTAAA--- | 6079 |
| Qy | 1002 | -----ProGlnAsnLeu----- | 1278 | LeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer | 1297 |
| Db | 5087 | ACTGAGCTGCTGAGCCACCTGAGGTCGCGGTGCGGTCGCCAAGGTCGCCAGAA | 6080 | -----GAAGTGGAGTCACTCTTGTTCGAAAGACAAAAGGCGCGCAG--- | 6121 |
| Qy | 1012 | ProGlnProGlySerSerProArgGly-----LysSerArgSerPro | 1298 | SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArg | 1317 |
| Db | 5147 | ACACGTCCTGAGGCCACCTGAGGTGGCGCCCAATAGCGGTGAAGAGAGCTCCATG | 6122 | -----AAACACCCGATCAGCGCGCAGCGAAGAAC | 6151 |
| Qy | 1027 | AlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPro | 1318 | ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro | 1337 |
| Db | 5207 | GAACCAAGGCTGCTGAGGAGGAGGAGTGAACAGAAACCTGACAGAAAAGAT--- | 6152 | ACAAACAGAAAGTG-----GTGGCTCTCTGTA | 6178 |
| Qy | 1047 | ProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSer | 1338 | GluArgHisSerProHisLysLysGluGlnHisIleArgGly-----Ser | 1354 |
| Db | 5264 | ---GCTGCACACAGAAAAACCCCTGAAACCCGCGCTGTGAAGTTGTAGAGAAAAA | 6179 | GAGAGCATGTCTCCT-----GAATCCAAACAGCTCAAGGTGAGAGTCTCTGCT | 6226 |
| Qy | 1067 | ProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeu | 1355 | IleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGlu | 1374 |
| Db | 5321 | CGG---GCCCTGAAATAATCCAAATCA-----AAGAGA | 6227 | GCAAAATGAGGGGACACAGTACAGCAGCAGCCGAGCCGAGCCAG--- | 6271 |
| Qy | 1087 | GlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAspProPro | 1375 | AlaLysLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThr | 1394 |
| Db | 5354 | CGAAGATCTCGAAATCCAGGTTAGCAGTGGACAAATCTGCAAGTCTGAAAAAT--- | 6272 | GAAAGCAGAGTGAGAAACCCCATCTCCCTCCTCAGTCATGTACTTCTGACCTAAGC | 6331 |
| Qy | 1107 | ProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSer | 1395 | GluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeu | 1414 |
| Db | 5408 | -----GTGGATGCTGCTCAGTCCAGCGGGGCTGCAGCAGCGAGGAGGGGAA | 6332 | AAGATTCCCTCCACAGAG-----GAA | 6271 |
| Qy | 1127 | GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly | 1415 | ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg | 1434 |
| Db | 5462 | TCTGGGTG---GTGGCAGTCTCCCTGAGAAAGTGAGAGTCCCCAAAGAGGAGTGGT | 6350 | -----AATTCGTCCCAAGAAATCAGTGTGAGGAAAGGACT | 6385 |
| Qy | 1147 | ProValThrMetGlyLeuProLeuProMetAspPro-----LysLysLeu | 1435 | HisThr-----ProGluLeuProLeuAlaProArgProLeuLysGluGly | 1449 |
| Db | 5519 | TTATCATCCAGTTGAAAGTATCCAGTTGATCCAGACAGAAACCAAGAGAAAGAAC | 6386 | CAAACCAAGCATCTGTGCGCCAGACCTTCCCCACCTCCCGAGCA--- | 6433 |
| Qy | 1162 | AlaProPheSerGly-----ValLysGlnGlnLeu--- | 1450 | SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLys | 1469 |
| Db | 5579 | GTGTCTGCTCTGGCGCGTCCCGAGAGCCAGTTCAGCAGAGTGGAGTGGAG | 6434 | -----GCACCGGTGGATGAGAGCTCAAGCC----- | 6460 |
| Qy | 1172 | ----- | 1470 | LysHisAspValArgSerLeuIleGlySerProGlyArgThrPhePro-----ProVal | 1487 |
| Db | 5639 | CAGCGCGTGAACACATCGAAGCTCGTGAGGCGCTCTGCTCTGCTCTATTAAGGCA | 6461 | AGGTTACGGGTGCATTCCATCATTGAAAGTGACCGCGTGACCCCGCCAGCGATCCCAAGC | 6520 |
| Qy | 1173 | ---SerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlu | 1488 | HisProLeuAspValMetAlaAspAlaArgAlaLeuGluAlaCysTyrGluGluSer | 1507 |
| Db | 5699 | GATGCACCAAGGCGCTTCC-----CCAGAGGACAGGACAGGCTGCACACCAAGCA | 6521 | ATCCCATATACCACTGCTCTGTGTA----- | 6547 |
| Qy | 1192 | AlaSerValLeuArgGlyThrAlaLeuGlySerVal----- | 1508 | LeuLysSerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaPro | 1527 |
| Db | 5753 | AGTGAACAGAGTGGCTGGCGCATCGCTCCATCATCAATGACATTTCTGGGAGGCCA | 6548 | -----ACTGCAGCAAGCTCTCACCTCTGCTGCTCTCTGGG----- | 6583 |
| Qy | 1204 | -----ProGlyGlySerIleThrLysGlyIleProSer | 1528 | ValIleValProGluLeuGlyLysProArgGlnSerPro----- | 1540 |
| Db | | | 6584 | -----GGGATCCACACAGAGCCCGCCCTCTAAGGTGACAGAGTGG | 6625 |

CC whether the comparison indicates a likelihood disease). (M1) is useful
 CC for identifying modulators of the p53 pathway. A probe for HM expression
 CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
 CC in a patient, where the cancer has greater than 25 % expression level.
 CC Modulators identified by (M1) are useful in a variety of diagnostic and
 CC therapeutic applications, where disease or disorder prognosis is related
 CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
 CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
 CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
 CC the p53 function of the cell, so that the cell undergoes normal
 CC proliferation or progression through the cell cycle. (M2) and (M3) are
 CC also useful for treating defects in the p53 pathway such as angiogenic,
 CC apoptotic or cell proliferation disorders. The present sequence is an HM
 CC nucleic acid encoding a p53 pathway modifying protein
 XX
 SQ Sequence 11145 BP; 3179 A; 2966 C; 2736 G; 2264 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,79e-09 Length: 11145
 Score: 586.00 Matches: 638
 Percent Similarity: 30.91% Conservative: 379
 Best Local Similarity: 19.39% Mismatches: 1071
 Query Match: 4.43% Indels: 1207
 DB: 7 Gaps: 150

US-09-522-753-5 (1-2517) x ACD13385 (1-11145)

QY 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleGln 60
 DB 995 GACTATAACCAAGATCGTACATATATGAGAGTGTGCAAGTCCAGGCACATTACCTGAG 1054
 QY 61 ProGlnArgArg---ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer 79
 DB 1055 GATTCAGCGCGGACTATCCAGCTCGAGGAGAGAGTTTATTTCAGAAATGGAAACTTAC 1114
 QY 80 Gln----- 80
 DB 1115 CAAGGAGACTACTATGAATACGATACGATGATCTCGGAATACAGGNTTACAGG 1174
 QY 81 -----GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96
 DB 1175 AATGATCCTTATGAACAAGATATTAGG-----GAATATAGTTACAGGCAAAAGGAACGA 1228
 QY 97 LysSerGluMetGluPheIleGluSerLysArgProArg-----Leu 110
 DB 1229 GAAAGAGAAACGTGAAGATTGAGTCTGACCGGACAGAGCATTGAGAGGAGCGCGATT 1288
 QY 111 GluLeuLeuProAspPro-----LeuLeuArgPro-----SerProLeu 123
 DB 1289 GAACGAGTCAAGTCTGTTTCACTTGCAGCTCCACAGAGTCTCGAGCGTCTCCCTCT 1348
 QY 124 LeuAlaThrGlnProAlaGlySerGlu-----AspLeuThrLysAspArg 139
 DB 1349 CAGGACAGAGAGGTGCGGAGTGTGATCTGAGAGAGAGGCTTTACAGCGGATCCTCAGACCGG 1408
 QY 140 SerLeuThrGlyLysLeuGluProValSerProPro----- 151
 DB 1409 AGT-----GGAAGCTGTAGCTCACTCTCCCTCCAGATATGAGAACTGGCAAGTCT 1462
 QY 152 -----SerProProHisThrAspProGluLeuValProPro 165
 DB 1463 CGTTTGAGCGCTATACAAAAATGAAAGACAGATAAAGAACGAACTTTTGATCCGGAG 1522
 QY 166 ArgLeuSerLysGluGluLeuGlnAsnMetAspArgValAspArgGluIleThrMet 185
 DB 1523 AGATGGAGAGAGAGACGCTTAATACGGAAGAAAAGTGAAGAAAGCAAACTGAC 1582
 QY 186 ValGluGlnGlnIleSerLysLeu-----LysLysLysGlnGlnLeuGluGlu 202
 DB 1583 AAGCAAAACGCAAGGAAGAGTTTCACTCCCTAGTTCTCAGTCTTTCAGAAACGGACCAA 1642
 QY 203 GluAlaAlaLysProProGluProGluLysProValSerProProIleGlu----- 220

DB 1643 GAAATATGAGCGAGAGCAAAAGCCCTGAAAAGCCAGGAGTTGTATAATAACTGAGCAGAGAG 1702
 QY 221 -----SerLysHisArg---SerLeuValGlnIleIleTyrAsp 232
 DB 1703 AAAGCTGACAAAGAGGGAATAGCGAAACCCCTGGAACTCATGCTGGTGGTGGTTTG 1762
 QY 233 GluAsnArgLysLysAlaGluAlaAala-----HisArgLleLeuGluGlyGly 249
 DB 1763 ACTCAGTGAAGAGAGAGAGGAAAGGATCATTCACACACTCTCTGTGGAAGTGAAG 1822
 QY 250 ProGlnValGlu-----LeuProLeuTyrAsnGlnProSer 261
 DB 1823 GCCAAGCTTGAATATGACACTGTCAAAATCTTCTGCCCTGGACCAAGAACTTCAGTCTCT 1882
 QY 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu--- 280
 DB 1883 CAGACGGAGCCTGCAAAATCTGACTTGTCTAAACTGGAATCAGTTAGATGAATGAATACCA 1942
 QY 281 -----IleLeuTyrPheLysArgArgAsnHisAla 290
 DB 1943 AAGGAAAGGGGCTTTCAAGCCATGTTGAAGTGTGGAGAGGAGGAGGCTTAAAGCC 2002
 QY 291 ArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu-----Met 306
 DB 2003 AGGAAGCCTCAAGCCCTGAG-----CAGCCTGAGATGGGTAAAGTCTGTGGATCTG 2056
 QY 307 GluAlaLeuGluLysLysValGluArgLleGluAsnAsnProArgArgArgAlaLysGlu 326
 DB 2057 GAGAACTGGAGCCAGG-----AAAAGGCGCTTTCAGAT 2092
 QY 327 SerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeu 346
 DB 2093 TCCAATTTAAAGCA-----GAAAGCAAAACCCAGAGGTCAAGAAA----- 2134
 QY 347 GlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAla 366
 DB 2135 -----AGCAGTCCAGAGATGGAGATGCT 2158
 QY 367 Arg-----SerGluHisGluValSerGluIleIleAsp 377
 DB 2159 CGCTGTCTTTCAAAAAGAGCAGCTGCTGCTCTAGAGAGGTCATTCGCTGAGGGAA 2218
 QY 378 GlyLeuSerGluGlnGluAsnLeuGluLysGlnMet----- 389
 DB 2219 CGAGAGCTGAAAGAAAGCCTGTGAGGAAGAAATCTTAAAGAGAGATCTAAATAATC 2278
 QY 390 -----ArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnArg 406
 DB 2279 AAAGTGGACAGACTTAATCTGTTGCCAGCCCT-----AAAGACTGTTCAGGAG 2326
 QY 407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys----- 424
 DB 2327 CTTGCCAGTATTTCTGTG-----GGTCTGCTCAAGGCCAGCTCAGACCTACAGCAAGAC 2384
 QY 425 ---AspArgGlnValMetAsnMetTyr-----SerGluGlnGluLysGluThrPhe 440
 DB 2385 TGGGAGAACTAGCAGGTGAATCTGTGGAAAATCAAGAAGTCCAATCAAAAAGGCCATTC 2444
 QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
 DB 2445 CCTCAAAACACACAGCTCAACAGCTGCAGGTATTAGATGATCAAGGACAGAGAGAGAG 2504
 QY 461 -----LysThrValAlaGluCysValLeuTyrTyrTyr----- 471
 DB 2505 AGTTAGGAAAACTATTTCAGTCTTCTGATGAACACACCTGAAACCTGAAATCAGGCCAAG 2564
 QY 472 -----LeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArg 485
 DB 2565 AGAAATCACATTCAGTAAATACTGAAGAAAAAATTTGGCAATTGCATTCATCAGCGCAGA 2624
 QY 486 SerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 505
 DB 2625 GT-TACCGAAAAACAAATGGACAGAGTCTGTAGGAACAGCAGATCGAAATGGAATAGCC 2683

| | | |
|------|---|------|
| 4486 | CTTGAAAAGTTGGAAAGAGGCTCTCCATGAGTACAGAGAAGACTCCAACGATAGAAATC | 4544 |
| Qy | 565 AlaserylsglyArglysthrAlaasnserelnglyargarglysgly----- | 580 |
| Db | 565 AlaserylsglyArglysthrAlaasnserelnglyargarglysgly----- | 580 |
| 4546 | AGCGAGTGCTCTCTGACAGCTTCAACGACCGAGGCGAGGCGGGGCTGACTGCCTTC | 4605 |
| Qy | 581 -----ArglleThrArgserMetalaasnGluAlaasnserGluGluAlaalleThr | 597 |
| Db | 581 -----ArglleThrArgserMetalaasnGluAlaasnserGluGluAlaalleThr | 597 |
| 4606 | CTGAGAGTCTCTTTACCGGAGCCACCTCGGAGACGACCAAGCGGAGGAGAGCGCTGCCTC | 4665 |
| Qy | 598 ProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTprThrGlu | 617 |
| Db | 598 ProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTprThrGlu | 617 |
| 4666 | CCTGAGAGCTGAAAGAGAGAGGAGAGGCACAGACCTCTCATCTTTCATCCAGAAGAGC | 4725 |
| Qy | 618 GluGluMetGluThrAlaLysLysGlyLeuLeuGlu-HisGlyArgAsnTrpSerAlaI | 637 |
| Db | 618 GluGluMetGluThrAlaLysLysGlyLeuLeuGlu-HisGlyArgAsnTrpSerAlaI | 637 |
| 4726 | CACGACCGAGCGAGCCCAAGAAAGAGAGAGCGCCGAGAAAGAGAGAGGCGGAGATTAC | 4785 |
| Qy | 637 eAlaArgMetValGlyserIysThrValserGlnCysLys----- | 650 |
| Db | 637 eAlaArgMetValGlyserIysThrValserGlnCysLys----- | 650 |
| 4786 | AAGGAGGCGGTAGCAGCAAGAGGACTCGCGGCAGTACGAAAGAGGACTTCTCGAGGCGGAT | 4845 |
| Qy | 651 -----AspPheTyrPheAsnTyrLysLys-----ArgGlnAsnLeuAspGluLleLe | 666 |
| Db | 651 -----AspPheTyrPheAsnTyrLysLys-----ArgGlnAsnLeuAspGluLleLe | 666 |
| 4846 | GCTTACCGAGGTTCTT-TACAACTGAAGCTGCATAGAGATGAGCTAGATAAAACCAT | 4904 |
| Qy | 666 uGln-----GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgly | 682 |
| Db | 666 uGln-----GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgly | 682 |
| 4905 | TGAATTGTTTCTACCGAAAGAAAGATAAATGATTCGGAGAGAGAACTTCCAGAA | 4964 |
| Qy | 682 sLysLysLysAlaProAlaAlaLaserGluGluAlaAlaPheProProValValGluAs | 702 |
| Db | 682 sLysLysLysAlaProAlaAlaLaserGluGluAlaAlaPheProProValValGluAs | 702 |
| 4965 | AATAGAAAAGAACTAAAGCCTTATGCTTAGTGCCATCAA-CATCCTAAAGAGAGA | 5023 |
| Qy | 702 pGluGluMetGluAlaSerGlyValserGlyAsnGluGluGluMetValGluGluAlaGl | 722 |
| Db | 702 pGluGluMetGluAlaSerGlyValserGlyAsnGluGluGluMetValGluGluAlaGl | 722 |
| 5024 | AGAAAGAGAGAAACACAGGCGAGAAATGGAGAGACGAGAA----- | 5063 |
| Qy | 722 uAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVa | 742 |
| Db | 722 uAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVa | 742 |
| 5064 | -----GGAGAGGCGACCGGGACAGCATCGGATGGGCTGCTGGCGC | 5104 |
| Qy | 742 lAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaLysAspTh | 762 |
| Db | 742 lAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaLysAspTh | 762 |
| 5105 | A-----TCACGGGACGAGCTCTCGGCGCATCACAGGCGAGCAGCAAGCCCGCCACCA | 5158 |
| Qy | 762 rGlylAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPr | 782 |
| Db | 762 rGlylAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPr | 782 |
| 5159 | GGGACAGGACAGCCCCCGCGTGTCTCAAAGACAAGTCCAGGAGCGAGGGCCCGA-GG | 5217 |
| Qy | 782 oProThrProPro-----ArgArgThrSerArgAlaProIleGluPro-----Th | 797 |
| Db | 782 oProThrProPro-----ArgArgThrSerArgAlaProIleGluPro-----Th | 797 |
| 5218 | CTCGGCGATGCCAACTGAAGGAGAAATTCAAGACCGGT-CGAGAGAAAGAAAGGGCGGA | 5276 |
| Qy | 797 rProLaserGluAlaThrGly-----AlaProThrProProPro----- | 810 |
| Db | 797 rProLaserGluAlaThrGly-----AlaProThrProProPro----- | 810 |
| 5277 | CCCAAGTGAAGTAGCAACCGGATGATAAGGTAGCGCCATCCAAAGACCCAGGCAAGAA | 5336 |
| Qy | 811 -----AlaProProSerProSerAlaProProProValProLysGluGluLysGluGl | 829 |
| Db | 811 -----AlaProProSerProSerAlaProProProValProLysGluGluLysGluGl | 829 |
| 5337 | AGACGCCAGGCCCAAGGGA-GAAAGTCTCTGGGGAGCGGCGACCTGATGACCACTTCG | 5395 |
| Qy | 829 uGluThrAlaAlaAlaProProValGluGluGluGluGluGluGluGluGluGluGluGlu | 849 |
| Db | 829 uGluThrAlaAlaAlaProProValGluGluGluGluGluGluGluGluGluGluGluGlu | 849 |
| 5396 | AGAGGATGCTGTC-----CCAGAAAGGACCTGGAGATCGAGGAGCGCCCAAGCGGCA | 5449 |
| Qy | 849 uGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGl | 869 |
| Db | 849 uGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGl | 869 |
| 5450 | AGGAGAGGATGAACCAATGGAGAGAGCTGAGCCACCGGTCCG----- | 5492 |
| Qy | 869 uAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGl | 889 |
| Db | 869 uAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGl | 889 |
| 5493 | -----AGACCCCAAGCTCAAGGAGAGG | 5515 |
| Qy | 889 uAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGl | 909 |
| Db | 889 uAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGl | 909 |

889 yAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSe 909
5516 CGAAGCGGCAGACGACGGCGGAAGAGGCTCGACATTCCTGCTTAAGAAACCGCCGG 5575
909 rGlyAlaPro-----GlnAspSerAspSerSerAlaThrCys-S 922
5576 GGCTGGACCTCCATTTAAAGACAAAAAGCTCAAAGAGTCGACTCTATTCCACCTGCCG 5635
922 exAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeu----- 938
5636 CGGAAATAGCTACACCCAGCATCAGGTCGACAGCTCAAAGACTGCGCTGGCAGGCCCTC 5695
939 -----SerProArgProSerLeuLeuThrProThrGlyAsp 951
5696 ACATGAAGAGGTCCTCGCTGCCCGAGGCTGACAGAGCGGCCCACTGGCGTGC 5755
951 roArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLys---GlnArgA 970
5756 CC-----ACCCCTACGTCGGTGCTATC-CTGCCCCAGCTACGAGGAGGTGATG 5802
970 laAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAla 990
5803 CACAGCGCCAGGACCCCGCTCTCGACGCCGATGACTACCGGACCTCGGT-TCGACTG 5861
990 laProThrLys-----ProAlaProAlaProProProPro- 1002
5862 CGCGACTCGCAGCATCTCACGCCCGTGCACCGCTCCACACGAGCGCTGCTCCCCCTC 5921
1002 ----- 1002
5922 CTTTTTCAGACAGTCTTCCTCGTGCTTCAAGTGGGCTTTCGGAACACCCAGCAGGCTCC 5981
1003 --GlnAsnLeuGlnProGluSerAspAlaProGln-----ProGlySerSerp 1019
5982 TGCAGGCTCTCTCACAAACCTTTACCGCTCGGTCTGTGTCACATTAGGAGGACCCC 6041
1019 roArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPhe---AlaAlaG 1038
6042 CGAGGAAGAATTTCAGCGTCGGAGACAAGCTCTTCAGGCAGCAGAGCGTCTCTGTGCTC 6101
1038 luAlaGlnLysLeuProGlyAspProProCysThrThrSerGlyLeuProPheProValP 1058
6102 CAGCTACGACTCTCCCATGCCCACCTCGA---TGAAGACAGAGCGGCCCTGCCCCCGGT 6158
1058 roProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyra 1078
6159 TCCGCGGAGAGTTTGCCTGTCTGTGCGCAGGGTACTACTCCCGACGACTATGGCTCCC 6218
1078 laProPro-----GlyHisProLeuProLeuGlyLeu-HisAspThr 1091
6219 GTCGCCCAAAGTCGAGCGTTTGCAGCTGCCACCGGCTCGCGTTGTCACTGTCAACCCGCTC 6278
1092 AlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSer 1111
6279 TCCAGAGGGCGCTCTTCTCAAG-----TTTTCAAGC 6308
1112 AlaLysHisProSerValLeuGluArgGlnIleGlyAlaIle---SergInglyMetSer 1130
6309 AAAACCTTCCCTTCCCCAGAGCCGAGCTGCTGTTCTTCCCTCGTCAAGAGGGCCCTTCC 6368
1131 ValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly-ProValThrMe 1150
6369 -----CCCGGACCTGGACACCTCCGAGGACCAAGCAGCGCAGCGCCGCGCAT 6413
1150 tGlyLeuProLeuPro-----MetAspProLysLysLeuAlaProPheSerGlyVally 1168
6414 CATCCCCCGGAGCCAGCTACTCTGGAGCGCTGGAGGGTCCCTTTCAGCCCGCTCAT 6473
1168 sGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProTh 1188
6474 CACCGAGGAG-----CCCGTTGAGTGGGCCACCCCTCCGACGAG----- 6513

(WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX Callen DF, Whitmore SA, Kremmidiotis G, Kochetkova M, Crawford J;
XX WPI; 2001-316439/33.
XX New nucleic acid representing the human tumor suppressor gene TSG16,
XX useful e.g. for diagnosis and treatment of tumors, inflammatory and
XX immunological disorders.
XX Claim 112; Page 167-75; 215pp; English.
XX The present invention relates to human tumour suppressor gene 16 (TSG16;
XX see AAH23688). The present sequence is a partial TSG16 genomic sequence.
XX TSG16 was isolated from chromosome 16q24.3. TSG16 suppresses cellular
XX proliferation. TSG16 is useful for treating disorders associated with
XX decreased expression or activity of TSG16, e.g. cancers, (auto)immune
XX disorders, inflammation, complications of wound healing and infections
XX (by viruses, bacteria, fungi, parasites, protozoa or helminths)
XX SQ Sequence 9115 BP; 2368 A; 2674 C; 2571 G; 1502 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9,05e-10 Length: 9115
Score: 593.50 Matches: 565
Percent Similarity: 33.49% Conservative: 266
Best Local Similarity: 22.77% Mismatches: 921
Query Match: 4.49% Indels: 741
DB: 4 Gaps: 113

US-09-522-753-5 (1-2517) x AAH23695 (1-9115)
QY 62 GlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGlu 81
DB 3008 CAAAGAGAGGAGAGAGATTTTAAAGAGAGATATAAGAAAAATCAAAAAGAGAAAGG--- 3064
QY 82 LeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLys- 97
DB 3065 -----TTTATAGGGAAGATTCGTCTTTG 3088
QY 98 -----SerGluMetGluPheLeuGluSerLysArgProArgLeuGlu----- 111
DB 3089 ACGAATATTGTAACAAAATCAGTTCTGGAGAAATCAAGACACCAAAATTTAGCCCTTTCTG 3148
QY 112 -----LeuLeuProAspProLeuLeuArgProSerProLeuLeu 124
DB 3149 ACGATCAGCGAGATCGGTGTTTCTGACTTGTCCGATT-----CATCCTTTGATT 3199
QY 125 AlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys 144
DB 3200 TCAAAGGGAGGAGACAGCTGGGACTCCGCA-GTGACAGACTACAGGGACATG----- 3249
QY 145 LeuGluProValSerProSerProHisThrAspProGluLeuLeuValPro 164
DB 3250 -----AAGAGCGACTCTGTGCCCAAGCTCATCTTG 3279
QY 165 ProArgLeuSerLysGluGluLeuLeuGlnAsnMetAspArgValAspArgGluLeuThr 184
DB 3280 GAGACGGTGAAGGAGACACAGAGAGAGCGGGAGACCCCGCCCGGAGAGCGA 3339
QY 185 MetValGluGlnGlnLeuSerLysLeuLysLysGlnGlnGlnLeuGluGluAla 204
DB 3340 GACTACAGAGCGCTTCTCCGA---AAGAGGACAGGGAGCTATTGTGGATAAAACTCT 3396
QY 205 AlalysProProGluProGluLysProValSerProProIleGluSerLysHisArg 224
DB 3397 GAGAAGAGGAAGAGACAGACCGA-----AAGCATATAA 3429
QY 225 SerLeuValGlnIleLeuTyrAspGluAsnArgLysLysAlaGluAlaHisArgile 244
DB 3430 AGTGTCCCTACCTTCCGAAAGAGGACACAGAGAGAGAGAGTCCGACAG----- 3483
QY 245 LeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArg 264

DB 3484 ---GCCGGCGGGACAGAAAGCGCGCTCGAGAGCTGCAAGGCGCGGACGGCAGG 3540
QY 265 GlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuLeuTyrPhe 284
DB 3541 -----GCCAAGCCCGAGGAGCGCGCCCGGAGGAGCTG----- 3573
QY 285 LysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGln 304
DB 3574 ---AAGGAGTGTGGCTCGAGAGTGGCTTCAAGGACAAGTCCGACGGCGACTTTGGGAAG 3630
QY 305 LeuMetGluAlaLeuGlu-----LysLysValGluArgIleGluAsnAsn 319
DB 3631 GCCTGGAGCGCTGGAAACGGCACCCAGCAGAGAGAGGAGAAAGATGGCCCC 3690
QY 320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
DB 3691 GATAAGGAAGAGAGGAGAGACAAACACGAAAGATACAAAGAGAAATCCAGTGACAAG 3750
QY 340 IleArgLysGlnArgGluLeuGlnArgMetGln----- 351
DB 3751 GACAAAGTGAGAAATCAATCTCGAAATAATGTCAAGAGGACAAAGATTTGATAATGT 3810
QY 352 -----SerArgValGlyGln 356
DB 3811 TTTAAAGAGAAAAAGATACCAAGGAAAAACATAAGACACATCGGCAACAAAGAA 3870
QY 357 ArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluLeuIle 376
DB 3871 AGGAAGCGTCTCTCCACCAAGGGAAGAGAGAGAGAGGCTTTCCTCGGATCATC 3930
QY 377 ---AspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIle 395
DB 3931 TCAGAGAGACTTCTCAAAAAAAGATGACAAAGAGGCAAGAGAAAGAGTGTGTACATC 3990
QY 396 ProProMetLeuTyrAspAlaAspGlnArgIleLysPheIleAsnMetAsnGlyLeu 415
DB 3991 GCAGACATCTTTCAGATGAGTAGTGAGAC-----GACAGAGACAGCTGC 4035
QY 416 MetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGln 435
DB 4036 ATGGGGAGCGGTTCAAGATGGAGAGCGCCAGCGACTTTCGCGAGGACGGACGGCTCCAG 4095
QY 436 GluLysGluThrPheArgGluLysPhe-----Met 445
DB 4096 GAGAAGGAGGAGAGCGGAGCGCTATGCTCCGACAGACACAGGAGTCTTCGACAAG 4155
QY 446 GlnHisPro-----LysAsnPheGlyLeu 453
DB 4156 CAGCACCTTGAGGCGCAGAGACAGGAGCCAGGAGAGAGAGAGAGGACCGAGGGGCT 4215
QY 454 IleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLeuThr 473
DB 4216 GCCGACGCGGGGAGACAAAAA-----GAGAAAGTCTTTGAAAAGCACAAGAG 4266
QY 474 LysLysAsn-----GluAsnTyrLysSerLeuValArgArgSerTyrArgArg 489
DB 4267 AAGAAGGATAAAGAGTCCACAAAAAGTACAGAC-----AGG 4305
QY 490 ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
DB 4306 AAGGACAGAGCTCAGTGGACTCCACGCAAGATAGAAAAATAAACAAGAGTCCCCGAG 4365
QY 510 GlnProMetProArgSerSerGlnGluLysAspGlu-----LysGluLysGlu 526
DB 4366 AAGGCTGAAAAAGAGACGCTGCCGAGACAAAGCGCTAAAGCTAACTCAAGAGAGAGTCG 4425
QY 527 LysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeuLeu 546
DB 4426 GACAAGAGACATTCAAAGAGAGAGGAGAGTCTCGAGAGTCCGACGCGCGGAAAAAGCCCTG 4485
QY 547 LysGluLysThrAspAspThrSerGlyGluAsp-----AsnAspGluLysGluAlaVal 564

| | | | |
|------|----|---|------|
| 5159 | DB | CAGCAGGGGGAAGCAAAACGTGTCTTTTCCACGAGTGGCAGCTTGAACTTACCATGGAAA | 5218 |
| 1314 | QY | -----MetclucGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetG | 1332 |
| 5219 | DB | CCACGACAGCAAAAAGATGAAGGAGGAGGCTTCCAAAGCAGATGAAAAACTT--- | 5274 |
| 1332 | QY | lyArgAlaIleProPro-----GluArgHis | 1341 |
| 5275 | DB | --GAGCTGTAAAAACCATGCAGTGTGTCTTACCAGCACCGAGGATGGGAAGGCCAG | 5332 |
| 1341 | QY | erProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProA | 1361 |
| 5333 | DB | AGAAATCCCACTTA-----GGCAACAGGGCCCAAGGC----- | 5364 |
| 1361 | QY | rgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLys--- | 1379 |
| 5365 | DB | -----AGGAGGAGCTAGAAAATGTGTCCAGCC | 5392 |
| 1380 | QY | -----ArgGluGlyThrProProProProSerArgAspL | 1393 |
| 5393 | DB | TGTCTCTCAGATCCATCTGCACAAAGGAATAATCTCCCTCCCTCACCTGTATGTCTCC | 5452 |
| 1393 | QY | euthrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeu-----LysProA | 1410 |
| 5453 | DB | ATGGGCAGCCCGAGCTCAGAGCTTTGAAGGAATATCCCCAGCGCATCTGGCAAACTCTC | 5512 |
| 1410 | QY | laHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProA | 1430 |
| 5513 | DB | ACCGAAGAGGGTTGACCCCACTG-----ACGGTGAAGA | 5545 |
| 1430 | QY | rgGluGluLeuArgHisThrProGluLeuPro-----LeuAlaPro-----ArgProL | 1446 |
| 5546 | DB | AAGAAGATTCCAGGAACAACCTGATCTCCCTCTGGCACCTCCGAGCTCTCTGCCTC | 5605 |
| 1446 | QY | eulysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrT | 1466 |
| 5606 | DB | TGTCAGAA-----ACGTCTCCAGACACGCCAAGTCACAAGAAGT-----A | 5647 |
| 1466 | QY | hrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProP | 1486 |
| 5648 | DB | CGGACTCAAGAAG-----GTACTCGAGTTCCTCCCA | 5677 |
| 1486 | QY | roValHis-----ProLeuAspV | 1492 |
| 5678 | DB | GCCTCCACACCAACCACTAATGTCTGTTGTCTATTTAAACTATACATAAGCCAAATCA | 5737 |
| 1492 | QY | alMet-----AlaAspAlaArgAlaLeuGluArgAla-----CysTyrGluGluSerLeuL | 1509 |
| 5738 | DB | TCCAGCATGCAATAGGAGGTCCTCTGTTACGCTGTTGTCGATGAATGTTGTACAACC | 5797 |
| 1509 | QY | ysSerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGly----- | 1525 |
| 5798 | DB | CCAACCTCCGGGGTTTCCACTAAAGCTGCTTGTCTCCCTCTCAGGCTTAAGCAGAAG | 5857 |
| 1526 | QY | -----AlaProValIleValProGluLeuGly----- | 1534 |
| 5858 | DB | TCAGCAAAAGACATACACCATTGCCACAGCTCCGCATCTCTGAGCGCAGGAGGCTTGTGC | 5917 |
| 1535 | QY | -----LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProP | 1550 |
| 5918 | DB | CATCCAGCTCCGGCAAGCCCCCGCATGACAGAGGTT----- | 5952 |
| 1550 | QY | heAlaGlyHisLeuPro----- | 1555 |
| 5953 | DB | -----CACCTCCCTTCACTGGTTTCCCGGAAGGCCAAGATCTAGCTAGAGTGG | 6004 |
| 1556 | QY | -----ArgGlySerProValThrMetArgGluProThrProArgLeuG | 1570 |
| 6005 | DB | AGAAGGAAGAAGAGGAGGGGAGCG-----GAGGAGGATGCTCTCGCTCCCTCC | 6055 |
| 1570 | QY | In-GluGlySerLeuSerSerLysAlaSerGlnAspArg-LysLeu----- | 1585 |
| 6056 | DB | AGAGAGGGGAGCGCGCAGGATCAAAATCTTCGAAGAGGGGTACAAATCAACAGACAGT | 6115 |

| | | | |
|------|----|---|------|
| 1586 | Qy | -----ThrSerThrProArgGluIleAla----- | 1593 |
| 6116 | Db | ATGTATATGTCGAGCGCGCGGAGGAATAATGTTGTGAGGAGTGTGGAAATTCGCT | 6175 |
| 1594 | Qy | ----LysSerProHisHisSerThrValProGluHisHisProHisProIleSerProTyrG1 | 1612 |
| 6176 | Db | GCAGAAGCCAGCATGCT-----GAAGAACAATCCGACCCA----- | 6215 |
| 1612 | Qy | uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh | 1632 |
| 6216 | Db | -CACTG-ACGTCCGCG-----CCTATGTGTGC | 6240 |
| 1632 | Qy | eAspProThrSerIle-----ProArgIleProLeuAspAlaAlaAlaIaTy | 1649 |
| 6241 | Db | AAGCACTGTCATTTGCTTTTAAACCAAGGGAATCTGACTAAGCATCAAGTCCA-- | 6298 |
| 1649 | Qy | rTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrIle | 1669 |
| 6299 | Db | -----AGCCCAACAGCAAAAGTCCCAAGACAGAGGGTCTGGAGAGCTGGAA | 6348 |
| 1669 | Qy | uIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGln----- | 1683 |
| 6349 | Db | GCCGAAGAAGAACCACTGACCACTGTTTCCAGGACTCGGAAGACGAGAGGGTTCAGAG | 6408 |
| 1684 | Qy | -----ThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnTh | 1699 |
| 6409 | Db | GCTGTGAGAGAGCACATGTTTTCGGACTGGAGGACTCGGACTCAGACTCAGACTGGAC | 6468 |
| 1699 | Qy | rAlaThr-----AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro--Ar | 1716 |
| 6469 | Db | GAAGACGAGGATGAGGATGAGGAGAGACCGAGGATGAGCTGTCCAGACCATCTCTCAGAG | 6528 |
| 1716 | Qy | gGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerG1 | 1736 |
| 6529 | Db | CGCGCCCGCGCTGCGCCACCATGCACTGCGGG-----CA | 6564 |
| 1736 | Qy | nValProHis-----LeuProValLeuValProThrProGlyThrPr | 1751 |
| 6565 | Db | GACTCTCTACCCATCTCTGGCCCTCAGGCCCAAGATGCTCCCGCTCTGGCAGCGAGGCC | 6624 |
| 1751 | Qy | o-----AlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSe | 1769 |
| 6625 | Db | ACAGGAGCACTCGGTCTCGGAAGCTGAGCGCTCAGACGACGAGAGCTGCTCCATGTCC | 6684 |
| 1769 | Qy | rSerArg-HisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrIlysProT | 1789 |
| 6685 | Db | AGCCAGACCATGCGGGCCTCCCTGGCTGGGACCGGCCCTCTGGGCTCTGT----- | 6737 |
| 1789 | Qy | hrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgG | 1809 |
| 6738 | Db | -----GGAG | 6741 |
| 1809 | Qy | luArgGluLys-SerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArg | 1828 |
| 6742 | Db | AAAGACACAGGCTGAGCTTGAAGTACACGCTGTGTCCCAAGAGACCG--TGTGTC | 6798 |
| 1829 | Qy | ProGlyThrGluGlnSerSerGlySerSerGlySerGlyGlyGlyGlySerSer | 1848 |
| 6799 | Db | CCAAGCAAAAGAA-----GCAGCGACG | 6819 |
| 1849 | Qy | SerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerPro----- | 1865 |
| 6820 | Db | CGTCCACCACTAGCCCGGAAACACTCGCTAACCAAAACGACTCATCTCCCGAGCATGC | 6879 |
| 1866 | Qy | -----ArgThrGlnAspAlaGlnGlnArgProSerValLeuHisAsnThrGly | 1882 |
| 6880 | Db | TCCCGGCGCGAGAACCAACGAGCTCAGCCCCCAAGCCCACTGCGCTGCAC----- | 6930 |
| 1883 | Qy | MetLysGlyIleIleThrAlaValGluProSerIlys-----ProThrValLeuArg | 1899 |
| 6931 | Db | -----GTGACCCAGAGAGGGCATGCGCCCTCTCCCTCTGTGGG | 6969 |

| | | | |
|----|------|---|------|
| Qy | 762 | rGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyPr | 782 |
| Db | 3180 | G-----CCCTCTTGGAGGACCCACGCGCTTCAGCAAGGAC | 3218 |
| Qy | 782 | oProThrPro-----ProArgThrSerArg----- | 791 |
| Db | 3219 | TCCACACCGAGGTCCGGTTCAGATCAGGAAGGAGAGAAACACGTCCTCAAGAAAT | 3278 |
| Qy | 792 | -AlaProIleGluProThrProAlaSerGluAlaThrGlyAla----- | 805 |
| Db | 3279 | TTCTGTATCCAGCACACCGACTCTCTTTGAGAAATCTGATCTCTCGAGCAGCGAGTGG | 3338 |
| Qy | 806 | -----ProThrProProAlaPro----- | 812 |
| Db | 3339 | CTTGGAAAGGGGAACACAAACCTCTGGCCACAGTTCCCATCACCCACCCTGCCCCACACGG | 3398 |
| Qy | 813 | -----ProSerProSerAlaProPr | 819 |
| Db | 3399 | ACGCTCTGCTCACTCCTCGAGCCTAAGTTGGTCCGCCAGCCCAACATTCAGGTTCTCTGA | 3458 |
| Qy | 819 | oProValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluG1 | 839 |
| Db | 3459 | GATCTAGTAATCTGAGGAGCTGACCGCGGACACAGAGCCAGCGCGCCCTTAAGGA | 3518 |
| Qy | 839 | uGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaG1 | 859 |
| Db | 3519 | ACCTGAGAAG-----ACTGAGGAGTTCCA | 3542 |
| Qy | 859 | uGluProValLysSerGluCysThrGluGlu-----AlaGluGluGlyProAlaLysG1 | 877 |
| Db | 3543 | ATGCCCCCGCAGCCAGCACACTTGCCTCCAGCTCCCGAGCTGAGAAAGCTCCACCCCAAAA | 3602 |
| Qy | 877 | YLYsAsp-----AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaG1 | 894 |
| Db | 3603 | GAAGAGTTGGCCCTGCGAGAGATGGCCCAATCATCAGGGAGTCCAGCTTCGAGTCCCT | 3662 |
| Qy | 894 | uLYsLysGluGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAs | 914 |
| Db | 3663 | TGTGCCT-----CTGCTCGCAGCCGAGCAGGAAAGCAATGCTCTTTAGTGG | 3713 |
| Qy | 914 | pSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyAspLy | 934 |
| Db | 3714 | GTCCAGCGCTCAGCCTCGTTTGTAGAGGGATGACCATGGGAAGCCGAGCCCGCATCC | 3773 |
| Qy | 934 | sAsnArgLeuLeuSerProArgPro-----SerLeuLeuThrProThrGlyAs | 950 |
| Db | 3774 | CTCATCTGACATGGCCGCCAAACCCCTGGGCACCCACCATGTTGACTGTCCCGAGCCACCA | 3833 |
| Qy | 950 | pPro-----ArgAlaAsnAlaSerProGlnLysPro----- | 960 |
| Db | 3834 | CCCAATGCCGAGAGATGCGGAGTTCAGCTTCAGAGCAGAGCCCAACGTTTCCCATTC | 3893 |
| Qy | 961 | -LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaProProIleGlnValThrLy | 980 |
| Db | 3894 | TGCCCATGATGACCGAGACACGACCAATCCTTTGACTATGGCAGCTGTGCTTGACA-- | 3951 |
| Qy | 980 | sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProAlaProPr | 1000 |
| Db | 3952 | -----GGCCCTTCTGCTCCAGCCCGAGTGGCTCCACC | 3983 |
| Qy | 1000 | o-----ProProGlnAsnLeuGln----- | 1006 |
| Db | 3984 | AGCCGGGGAGGCCCGCCGACAGAGAAGAAATGCTTCTTGTGTGAGAGCCCTCTCTGTAG | 4043 |
| Qy | 1007 | -----ProLysSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerAr | 1024 |
| Db | 4044 | CAGGCTCCAGAACTCTGAGTTGGAGTTGGCCCC-----AAGGGAAGACAGGAGCGGA | 4097 |
| Qy | 1024 | gSerProAlaProProAlaAspLysGluAlaPhe-AlaAlaGluAlaGlnLysLeuProG | 1044 |
| Db | 4098 | AGAACCCAGCGCTCATCCAGTAACCCCTCTGCCAAAAGCTCATTTGTCCAGGATTTCTCT | 4157 |
| Qy | 1044 | lyAspProPro----- | 1047 |

| | | |
|----|---|------|
| | | |
| Db | 4158 TGGGCGCCACTTCACATGGTGGACCCCCCGGAGGCAAGGGGCCAGAGGCAGACAGGCCCGC | 4217 |
| Qy | 1048 -----CysThrThrSerGlyLeuPro- | 1054 |
| Db | 4218 ATTGGGGCCCATCTGTGCCTTACAGAAGCACATGCAAGTGTTCCACCACCCCGTTGCCCA | 4277 |
| Qy | 1055 -----PheProValProProArgGluV | 1062 |
| Db | 4278 GACACCCCTGCGATGAGAAAGCCATACCTGCCCCCACCACCATGCTCCCTTTCTCTCTCCACGA | 4337 |
| Qy | 1062 allle::: -----LysAlaSerProHisAlaProAspProS | 1073 |
| Db | 4338 TCTCGTGCAGCATGAGCCAGACAGTCTCAGAAATTCTTCTCCACCAGGCCCATGTCCAG | 4397 |
| Qy | 1073 erAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaA | 1093 |
| Db | 4398 CTTCTGTCTCACCATCTCCATG-CCCCCACTTCTCTCCCTCTTATTTCNAGCC---- | 4452 |
| Qy | 1093 rgProValLeuProArqProProThrIleserAenPro----- | 1105 |
| Db | 4453 --CCACCGCTTCTCTCTCAGCTACTGTCTGCACCCAGGCCAGTCCATCTCSCCCAGC | 4510 |
| Qy | 1106 -----ProProLeuIleserSerAlaIylHisProSerValLeuGluArgGlnI | 1122 |
| Db | 4511 TCATGCTCACCCAGCCAACATCCCCTTCAGGCAACCCCTTCTCTCTCCCATGCCAT | 4570 |
| Qy | 1122 le-----GlyAlaIleserGlnGlyMet----- | 1129 |
| Db | 4571 ACCGAGCTCTCAGCACTGTCTCTGGGTTTTCTGCCTCTGCAATCCAGTTTGCAC | 4630 |
| Qy | 1130 -----SerValGlnLeuHisValProTyrSerGluHialalysalaProv | 1145 |
| Db | 4631 TTCAGCTCCCTGGTGATGTGGAAGCCATCTGCCCCAGATCAAACACAGCTGSCCCAC | 4690 |
| Qy | 1145 alGlyProValThrMetGlyLeuProLeuProMetAspProLysLeuAlaIylProPheS | 1165 |
| Db | 4691 TGGCAACAGGAAGTGTGGC-----CTCTCCCCCAGCC | 4723 |
| Qy | 1165 erGlyValLysGlnGlu--GlnLeuSerProArgGlyGlnAlaGlyProProGluSerL | 1184 |
| Db | 4724 AAGAGTACAGCAGTGACATCGGTCTACCCCT-----GTGGCTCCCCCAGCCAGCT | 4774 |
| Qy | 1184 euGlyValProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeuGlySerValP | 1204 |
| Db | 4775 CCTCAGCACCTTACATCAGCTCCTCA-----CTGGCCCTGCCTGCCTGTC | 4819 |
| Qy | 1204 roGlyGlySerIleThrLysGlyIlePro-----SerThrArgValProSerAsps | 1221 |
| Db | 4820 CAGACACCATGTGTCTCCTGGTTGTGCTGTGTGCACCAATATGCCCGTCTATG | 4879 |
| Qy | 1221 erAlalleThrTyArgGly-----SerIleThrHisGlyThrProA | 1235 |
| Db | 4880 GGAGCGCAATGTACACACCTTTCCAGATCTTGGTCACCCAGTCCCAAGCAGCTCAG | 4939 |
| Qy | 1235 laAspValLeuTyR-----LysGlyThrIleThrArgIleI | 1247 |
| Db | 4940 CAACCTGTGGCACTTCCCAAGTTTGAGGAACCCCATCAAGGGG-----ACGACTGTAT | 4993 |
| Qy | 1247 leGlyGluAap-----SerProSerArgLeuAspArgGlyArgGluA | 1261 |
| Db | 4994 GTGGTGCAGATGTGCATGAGGTTGGGCGCCGCCCTTCTGGGTTAAGTGAAGAGCAAGACA | 5053 |
| Qy | 1261 spSerLeuProLysGlyHis-----ValIleTyrgluGlyLyalyesglyHisV | 1277 |
| Db | 5054 GAGCTTTTCCCACTTCCATCCTGAGATGCTGTGACATTAACCTGGAAGAAAAGCACTT | 5113 |
| Qy | 1277 alLeuSerTyrgluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerS | 1297 |
| Db | 5114 CCCTGTATCNAGAGAGTATCTTGAGCTG-----GAGGGGAGTTCATCAA | 5158 |
| Qy | 1297 erSerGlyProProHisGluThrAlaAlaProLysArgThrTyArgMet----- | 1313 |

Db 1414 ACAGAGAGGACACAAAGAGGAGGAGAGCCCGGAGAGCCAGGCAAGTACATCTGCCAG 1473
 Qy 117 euLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrL 137
 Db 1474 TACTGCAGCGCGCCCTGTGCAGAGCCAGCGTGTCCA-----G 1512
 Qy 137 ysAspArgSerLeuThrGlyLysLeuGlu-----ProValSerPro----- 150
 Db 1513 AAGCACATTGCTCACACACAGAGGTGAGAGGCGCCCTACCCCTCGCGCCCTGTGCTTCTCC 1572
 Qy 151 -----ProSerProHis----- 155
 Db 1573 TTCAAGACCAAGAGTAATCTTACAAGACAGGAGTCCCATGCCCHCCAGCATCAAGCA 1632
 Qy 156 -----Thr-AspProGluLeuLeuVal 163
 Db 1633 GGCCTGGCTCAGCATGGTGGCGAGATGTACCCACATGGCTGGAGATGGAGCGATC 1692
 Qy 164 ProProArgLeuSerLysGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 183
 Db 1693 CCTGGGGAAGAGTTTACAGAGCC-----ACTGAGGGGAGAAAGC 1731
 Qy 184 ThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnGlnGlu 203
 Db 1732 ACAGATTCTGAGAGGAGACTAGTGCACCTCTGTCTACCCCTGCAGAGCTTCC----- 1785
 Qy 204 AlaAlaLysProProGluProGluLysProValSerProProIleGluSer----- 221
 Db 1786 -----CCAGAGCCCAAGCAGAGCCCTCTCTCCAGCGGCTATACAGCTCTGGG 1833
 Qy 222 LysHisArgSerLeuValGlnIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 241
 Db 1834 AGCCACAGTCTCAGCAGAGCGTGTCTCCCTGTCCAGTCCAGCAGACCGCATCTACTC 1893
 Qy 242 HisArgIleLeuGluGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
 Db 1894 GAAGACCCCTCCATTGTGGAACCTCTACTGAGCAGCCCTGAGCCATAAACCTGAA 1953
 Qy 262 AspThrArgGlnTyrHisGlnLeuIleLysLysLysLysLysLysLysLysLysLysLys 281
 Db 1954 GACACCCACAGCATTAAGCAGAGAGTGGCCCTCCCTCTTAAGCAGGAGGAGAGGTGATC 2013
 Qy 282 -----LeuTyr-Phelys 285
 Db 2014 GATGACAGCGCTTCTGAGCCAGCAGGAGGAGTACTGAGTCTGGGTATTTCTCT 2073
 Qy 286 ArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGlnLeu 305
 Db 2074 CGCTCCGAGTGCAGAGCAGCAG----- 2097
 Qy 306 MetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAlaLys 325
 Db 2098 -----GTGAGCCCGCCCAACACCAACCCCAAG 2124
 Qy 326 GluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGlu 345
 Db 2125 -----TCCTACGCTGAGATCATCTTTGGCAAGT----- 2154
 Qy 346 LeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAla 365
 Db 2155 -----GGGCGAATAGACAGCGGAGCGCCATGTGTGACAGCCACTCC 2196
 Qy 366 AlaArg-----SerGluHisGluValSer----- 373
 Db 2197 ACCAGCCCTCTCTGCCCTCTCCAGGAGACAGCCAGCCAGCCCTGTGCTTGTCTGTA 2256
 Qy 374 -----GluIleLeuAspGlyLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 390
 Db 2257 CCCCAGCGAGGTGATCGAGCATCATCAGAGCTCATCACCATCAACGAG----- 2307
 Qy 391 GlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIle 410
 Db 2308 -----GCCGTGTGGACACCACTGATCGACAGCGTGAAGCCAGCGGAGCTCAGCTG 2361

Qy 411 AsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsn 430
 Db 2362 TCCAGCGCAGCAGATGGAGTCCCA---AAATCCAGCTCTACCGGAGCCCTGTGTA 2418
 Qy 431 MetTyrSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsn 450
 Db 2419 TCCACAGTGAAGAAACCAAGCCTGAACAATCACTGCTGAGCTCCAGCACCGCCAGT 2478
 Qy 451 PheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyr 470
 Db 2479 ACGGCCCCCTGTGCTCTCTCGAAGCCACTCAAT---GCCTTCTGCGCTGACAT 2537
 Qy 471 TyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArg 490
 Db 2538 CAGCACCCCCACCCCTTCGAGGTAGTACTCTCTCGATGACCATATCACCGACTC 2597
 Qy 491 GlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 510
 Db 2598 CGAAGCCCTGAGCGCAGCAGTCACGTGTTTACCTCCACCCCGGATGCTGAAGCCGCA 2657
 Qy 510 nPro-----MetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLys 527
 Db 2658 GCGGCAATCGAATTACCTTTGGGAGGAATACAGTTCTGAGAGCTGGCCCAAGCAG 2717
 Qy 527 sGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeuLeu 547
 Db 2718 CAAAGACACAGCTCCAGCCCTCGAGCAAGTGGAA---CCCAAGGAAGCGAGCTTAC 2774
 Qy 547 sGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer 567
 Db 2775 CAAAAGAGC----- 2789
 Qy 567 sGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly-----ArgIle 582
 Db 2790 GGGTTTGAACA-----AAAGGGTCTACTCGAATGTAACAT 2828
 Qy 582 eThrArgSerMetAlaAsnGluAlaAsnSerGluAlaIleThrProGlnGlnSerAl 602
 Db 2829 ATGTGGTCTCGGTACAGAAAGGATTAATACTAGAGGCCCAACAAATACTACTGCTC 2888
 Qy 602 aGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluGluMetGlu 622
 Db 2889 AGAGCTTCAGATCGCAAGGCCATCTCTGAGGAGCCACACATCTCCAGAA----- 2940
 Qy 622 rAlaLysLysGlyLeuLeuGluHisGlyArgAsnTyrSerAlaIleAlaArgMetVal 642
 Db 2941 -GCTGAAGAGAGTCAGATTGAGCAT---GAGCGGTGCTCCAA----- 2979
 Qy 642 ySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLe 662
 Db 2979 ----- 2979
 Qy 662 uAspGluIleLeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAlaArg 682
 Db 2980 -----ATGATGATTACAACTCGGACCCACCTGGAACCTACTCCAGTGGAGAA 3029
 Qy 682 sLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProValValGluAs 702
 Db 3030 GAGGAGGAAGAGAGAGCTTGGGAGCAGAGAGAG-----CCACCTGCTTTGATGC 3083
 Qy 702 pGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluAlaG 722
 Db 3084 CACAAAAGTCAGTTTGGCAGC----- 3105
 Qy 722 uAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVa 742
 Db 3106 -----CCCGGCGCATCTGTATGC 3122
 Qy 742 LasnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspTh 762
 Db 3123 TGCTCGGAACCTTCCCTCGAGTCCACCAAGTCCACCA---GCAGAAACCAAGTAATCAGT 3179

| | | | | | | | |
|----|-------|--|-------|----|-------|--|-------|
| Db | 14746 | AACTGCTCGGCGAGCGATCGCGCGGGAACCGGACCGGATCGCGACGATGCGGATG | 14687 | Db | 13703 | GGCCAGCACCAGCGCTCGGCTCCCGCTCGCGGAGC | 13647 |
| Qy | 1781 | -----GlyProThrHisLeuThrLysProThrThrThrSerSer | 1794 | Qy | 2050 | oGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLe | 2070 |
| Db | 14686 | GGCTCGGACTCGTCGGCGCGCGACGACGCGACCGCGCTGGTGGCGTCCGACG | 14627 | Db | 13646 | GACGGGGCTC-----GCCGCGCGGTGACGACG | 13602 |
| Qy | 1795 | GluArgGluArg-----Asp | 1799 | Qy | 2070 | uProLysHisLeuGluLeuAsp-LysSerHisLeuGluGlyGluLeu | 2086 |
| Db | 14626 | AGCGGGCGCGCAGGTGCTGCGCGAGCGCGGTGAGGTGGGTGGTGAACACACGCGAC | 14567 | Db | 13601 | C-----CAGTCCCGCAGCGGACTGCGCGAGGCTTCGCGGAGCAGCGACGCGCGCTG | 13548 |
| Qy | 1800 | ArgAspArgGluArgAsp-----ArgAspArgGluArgGluLysSerHisLeuThrSer | 1817 | Qy | 2087 | -----ArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeu | 2104 |
| Db | 14566 | CGGGCAGCGCAGCGCGGTGCTCGCGCACAGGCGCTTGGCGAGGCGCATCGCTTGAGC | 14507 | Db | 13547 | GAGCGCGCGTCCCGACCGCGAGGTGCTCTCGCACCGCGCGCTCGCGCGAGACTGC | 13488 |
| Qy | 1818 | -----ThrThrThrValGluHisAlaProIleTyr | 1827 | Qy | 2104 | roHisLeu-----ArgProLeuProGluS | 2112 |
| Db | 14506 | GAGTCGAGTCGAGCTCGCTGAACCGCGCGCGACGTGATCTCGCGGCGGACTCGTGG | 14447 | Db | 13487 | CACGAGCAGCAGCGCGCGCTGCTGACCGATTCGGGCGAGTGCACACA-CGCGCCC | 13429 |
| Qy | 1828 | ArgProGlyThrGluLysSerSerGlySerSerGlySerSer-----Gly | 1842 | Qy | 2112 | erGlnPro--SerSerSerProLeu----- | 2119 |
| Db | 14446 | ---CCAGCACTCGCGCACCGCGTTCGCGACGAGTTCGAGCAGGTTTCTCTGCTGCTCG | 14390 | Db | 13428 | AGCGCGCGCGAGCTCCAGCCCGCGGACCTGCGCGAAGCGGTGCGGTGCTCTGCGCAG | 13369 |
| Qy | 1843 | GlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerPro | 1862 | Qy | 2120 | -----LeuGlnThrAlaProGlyValLysGlyHisGlnA | 2131 |
| Db | 14389 | TCGGGGAAGCGCGCGCGTGTGTCGCGC-----TCGCGC | 14351 | Db | 13368 | GCGCGCGAGGTGCTGCTCGCGGGGAGTGCAGACCGCACCTCGGTGAGGACCAAC-C | 13310 |
| Qy | 1863 | IleSerPro-----ArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn | 1880 | Qy | 2131 | rgValValThr-----LeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrA | 2149 |
| Db | 14350 | GTCGCTCGCTCGCGCTCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG | 14291 | Db | 13309 | GGTGTGCACCCCGCATTCGCCAGCGCTCGGACCGAGGTGAGCAGCGCGCACCG | 13250 |
| Qy | 1881 | ThrGlyMetLysGlyIleThrAlaValGluProSerLysProThrValLeuArgSer | 1900 | Qy | 2149 | rgHisPro-----GlnGlnLeuSerAlaPro-- | 2158 |
| Db | 14290 | AGCGCGTGGCGCGGTG-----GCGCGAAGCCTTCTGCTGAGCACCGCGCAGTCCACG | 14237 | Db | 13249 | GCAGGAACCCCGCGAGCGCGGATTCGTCGACGACGACGACGACGCGCGCGG | 13190 |
| Qy | 1901 | ThrSerThrSerSerProValArgProAlaAla----- | 1911 | Qy | 2159 | LeuProAlaProLeuTyrSerPhe-ProGlyAlaSer-----CysPr | 2172 |
| Db | 14236 | TCGGCCACCGCACGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 14177 | Db | 13189 | AATCCCGTTCGCGCTGTCGCGCGCGCTGCGCGCATCTCAGTGGCGGAGGAGCGTGC | 13130 |
| Qy | 1912 | -----ThrPhePro | 1914 | Qy | 2172 | oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProPro----- | 2187 |
| Db | 14176 | CTGTGCGCGGAGGTGCGCAGTCCGCTCCCGCAGGTAGCGCTCGTCCACCGCGCG | 14117 | Db | 13129 | A-----CCTCGCAGCGCAACGCC-CTGCCACCGTTGGCTGTGATGGC | 13089 |
| Qy | 1915 | Pro-----AlaThr-HisCysProLeuGlyGlyThrLeuAspG | 1927 | Qy | 2188 | -----ProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyG | 2203 |
| Db | 14116 | CCCGCAGCGCCACCGCGCTCCAGCGCACCGAGGTGCACGAGCGCGCGCGCGCGCG | 14057 | Db | 13088 | CTCGCACGACATTCGCTCCACTCCCGACTGATGAGGGGTGACAAACCAACGAGTCGCC | 13029 |
| Qy | 1927 | yValTyrProThrLeuMetGluProValLeuProLysGluAlaProArgValAlaAr | 1947 | Qy | 2203 | yLysArgSerProGluProAsn-----LysTh | 2212 |
| Db | 14056 | TGTCGCGCCAGCGCTGAGGTAGCGCTGCTGCGCGCTAGCGCGCCATCCCGCGCGC | 13997 | Db | 13028 | GTCAGGTTTCCGACTCCCGCTTTCGATCGGAGTCCAGAAACGCGGTAGGCCAGCTC | 12969 |
| Qy | 1947 | gProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSe | 1967 | Qy | 2212 | rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSer----- | 2226 |
| Db | 13996 | CCCC--AGATCCCGCGACCGACGAGCAGTAGACCTCCGCTCGACGCGCGCGTGGCG | 13940 | Db | 12968 | GTCGAGGC-GGCGCGGTGCGCGCGGTGAGCCAGTACGCTCGCTGGAACGGAT | 12910 |
| Qy | 1967 | rgLysLeuLysProLysSerProSerLysGlySerGluProArgPro-----LeuVa | 1985 | Qy | 2227 | -----ProProGluGlyMetTh | 2232 |
| Db | 13939 | AGGACTTCGTCACGACGCGGACGCGG-----TCTTGCCCGCATGCTCTCGCGGAC | 13883 | Db | 12909 | ACGTGCGCAGCGCGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 12850 |
| Qy | 1985 | 1ProProVal-----SerGly--HisAlaThrIleAl | 1995 | Qy | 2232 | rgLupProGlyHisSerArgSerAlaValTyrProLeuTyrArgAspGly-GluGlnT | 2252 |
| Db | 13882 | TCCTCCGCTGCGACTTCGCTGATGCTGCGGAAGTTCGTGAGGGTTCGCGGTGACGAGC | 13823 | Db | 12849 | C-----GGACACGTGCGCGGTGCGCATCGACCG-----CAGGAAGCGCGAGGT | 12805 |
| Qy | 1995 | aArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspPro----- | 2011 | Qy | 2252 | hrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla--- | 2270 |
| Db | 13822 | GCGGTACCGCGTCTGTCGCGAGCGCTCGCGCAGCGCGCGCGCGCGCGCGCGCGCG | 13763 | Db | 12804 | CACCGCGCTCGCGGTGACGAGCGCGCATACGACCGCGTCTTCGGCAGCGCTCTT | 12745 |
| Qy | 2012 | -----ProAlaProProAlaSerAspProHisArgGluLysThrGlnSerLysPr | 2030 | Qy | 2271 | -----PhePheSerL | 2274 |
| Db | 13762 | ACGCGCGAT--CGGCGCGCTTCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 13704 | Db | 12744 | CGACCGCTCTCTGACGCGCTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 12685 |
| Qy | 2030 | oPheSerIleGlnLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr | 2050 | Qy | 2274 | ysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLysL | 2294 |
| | | | | Qy | 12684 | CGTATCCGACTCCCGCAGCGCGTACGCGGTGCGGAAAGCGGACCGCTCTCGCGCAGGT | 12625 |

| | | | |
|----|-------|--|-------|
| Qy | 1110 | rSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSe | 11130 |
| Db | 16499 | GTGCGCGCGCCACAGCTCGCCAGCAGACCATGATCGAGAAG---AGCACCGCGCTGGAG | 16443 |
| Qy | 11130 | rValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMe | 11150 |
| Db | 16442 | CAGCTGACGCGC-----TCGTGCGGGTCGGACCGCGCTCGCCCGCGGACAC | 16395 |
| Qy | 11150 | tGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGl | 11170 |
| Db | 16394 | GTCCAG-CAACTTCCAGTC-----GGTGTGGCGGAGA | 16363 |
| Qy | 11170 | uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGl | 11190 |
| Db | 16362 | CGCGCTC-----GGCGCACCGGCTCATCGACTCGCGG--AACACCTCGGA | 16320 |
| Qy | 11190 | nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValPro---GlyGlySerIleTh | 1209 |
| Db | 16319 | GGATTCAGCAGATCGGTGCCATGCCGACCCCATCGCGCGCTCGCGGGGAGACCGAG | 16260 |
| Qy | 1209 | rLysGlyIleProSerThrArgVal-----ProSe | 1219 |
| Db | 16259 | GACCGGCTTCGCGCGCGGAGGTCAACCGCGCAGCGCGTCCGCGCAGCGCGGCCCTC | 16200 |
| Qy | 1219 | rAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTy | 1239 |
| Db | 16199 | GGCCAGCGCG-----TCCAGCTCCGCGCAC--ACGCCGCGG----- | 16166 |
| Qy | 1239 | rLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyAr | 1259 |
| Db | 16165 | -----CGGTGCTGCCGACGACGCGGGCTCGCACGCTC | 16134 |
| Qy | 1259 | gGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSe | 1279 |
| Db | 16133 | GAACCGCGCCCTCGCGTCCGAGCGTC-----CACGCCGCGTC | 16095 |
| Qy | 1279 | rTyrGluGlyGlyMetSerValThr---GlnCysSer----- | 1290 |
| Db | 16094 | GGTCAGCCGAGCTCCCGGGTGTCCGCGAGGTCTCGCGCAGCGCGCGGCGCTCGGCCGC | 16035 |
| Qy | 1291 | -----LysGluAspGlyArgSerSerSerGlyProProHisGluThrAl | 1305 |
| Db | 16034 | GACCACGGCTTCCTCGCGCGGACAGCAGAACGCGCAGCAGCGCTCGTTCGCGGC | 15975 |
| Qy | 1305 | aAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyAlaIleSerSerAl | 1325 |
| Db | 15974 | GGCTCTCG-CTCTGTGGTCTCTTCACGATCACGTG-----CGGTGGTTCGCG | 15922 |
| Qy | 1325 | aSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLe | 1345 |
| Db | 15921 | TGACGCCGGAAGACACACCCCGCCCGCGGGCGCTCACCCCGCGGCACCGCACCT | 15862 |
| Qy | 1345 | uLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGl | 1365 |
| Db | 15861 | CCTCGCGCAG--CACTCCACGGC-----CCCCGACTCCACGTGCA | 15823 |
| Qy | 1365 | uAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArgGluGlyThrProPr | 1385 |
| Db | 15822 | T-----GTGCGGGACAGCTCGTCGCGGTGCAACACCGGGGCGACATCC | 15778 |
| Qy | 1385 | oProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLe | 1405 |
| Db | 15777 | CGTGCCGCATCGCCAGCAGCACCTTGATCACGCC----- | 15744 |
| Qy | 1405 | uLysLeuLysProAlaHisGluGlyLeu---ValAlaThrValLysGluAlaGlyArgse | 1424 |
| Db | 15743 | -----CGCGGCGCGCGCGGCTGGGTGGCGCATGTT-----GG | 15706 |
| Qy | 1424 | rIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProAr | 1444 |
| Db | 15705 | TCTTTCAC-----CGAGCCAGCGCAGCGCGCGGTCTCGCGTCT | 15667 |

| | | | |
|----|-------|---|-------|
| Qy | 1444 | gProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSer | 1464 |
| Db | 15666 | GGCGGTAGGTTCGCGATGAG---CGCGCCGGCTCGATCGGGTCGCGAGCTCGGTGCCCG | 15610 |
| Qy | 1464 | rThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPh | 1484 |
| Db | 15609 | TCCCGTGCGCCTCCACGAGCTCGACGTCGCGCGCGCGACGCCGAAAGCGCGAGCGCT | 15550 |
| Qy | 1484 | eProProValHisProLeu--- | 1491 |
| Db | 15549 | GCGCGAT--CACCGCTCTTCGACGCGCGCTTGGAGCGCGAGACCGTTGCTGGCAC | 15493 |
| Qy | 1491 | pValMetAlaAspAlaArgAlaLeuGluAArgAlaCysTyrGluGluSerLeuLysSerAr | 1511 |
| Db | 15492 | CGTCTGTGTGACCGCGCTGCGCGCAGCAGCGCCAGCACCTGGTGCCTGTGGCGCGCG | 15433 |
| Qy | 1511 | qProGly-ThrAlaSerSerGlyGlySerIle-----AlaArgGlyAlaP | 1527 |
| Db | 15432 | CCCGGAAGCGCTCCAGACACGCGCGCGCAGCGCTTCGACAGCGCAGACCGTGG | 15373 |
| Qy | 1527 | roValIleValProGluLeuGlyLysProArgGln---SerProLeuThrTyrGluAspH | 1546 |
| Db | 15372 | CCCGCGGAGAACGCTTGCAGCAGCCTCGAGCGCGCGCGCTCGCTGCTGAAGT | 15313 |
| Qy | 1546 | isGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProT | 1566 |
| Db | 15312 | CGACGAAGGTGTACGGGTTCGACATGACCGTCAGCGCGCGCGCAGCGCAGGAGCACT | 15253 |
| Qy | 1566 | hrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT | 1586 |
| Db | 15252 | CACCGCGCGCAGCGCTCGATCGGAGGTGATCGGACCGACGAGCAGCAGCAGCGCG | 15193 |
| Qy | 1586 | hrSerThr-----ProArgLuiLeAlaLysSerProHisSerT | 1599 |
| Db | 15192 | TGTCACCGTCAGCGCGCGCCCTCCACGCCGAGGTGTAGCGATGCGCCGCGAGACA | 15133 |
| Qy | 1599 | hrValProGluHisProHisProIleSerProTyrGluHisLeuArgGlyVals | 1619 |
| Db | 15132 | CGTCCGGAG--TTTCCGAGGCCCTGTAGCGCTC---GACCGCTC-----GCCTCA | 15083 |
| Qy | 1619 | erGlyValAspLeuTyrArgSerHisIleProLeuAla-PhAspProThr----- | 1635 |
| Db | 15082 | CCGCGCAGCAGCTG-----CATGTAGACTCGCGCTTCATCCGACGAGCAGCGCG | 15032 |
| Qy | 1636 | --SerIleProArg--GlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgH | 1654 |
| Db | 15031 | GTGTGCTGCTCGCAGGCGCTGGGTGCATGCCCGCGCTTCACCTGCCTCCACGCC | 14972 |
| Qy | 1654 | isLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrP | 1674 |
| Db | 14971 | GTCTCCAGCATGAGCG-----CTGCTCGG----- | 14946 |
| Qy | 1674 | roAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnG | 1694 |
| Db | 14945 | -----GTCCATCGCACGCTTCGCGGG----- | 14922 |
| Qy | 1694 | lnMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg--- | 1713 |
| Db | 14921 | -----CGTGATCGCGAAGACCGCGGTGAAATCCGCGCGCTCGGTG | 14879 |
| Qy | 1713 | euSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleA | 1733 |
| Db | 14878 | AGGAACCC-----GCCCTGTCCAGTAGTGT----- | 14850 |
| Qy | 1733 | spLeuSerGlnValProHisLeuProValProProProThrProGlyThrProAla | 1752 |
| Db | 14849 | -----GCCCGGGTGTCCGGTCCGGGTGAGAGCGCGCCGATGTCACGCGCCGCTCG | 14795 |
| Qy | 1753 | ThrAlaMetAspArgLeuAlaTyrLeuProThrAlaPProGlnProPheSerSerArgHis | 1772 |
| Db | 14794 | GCCGGGAAGCCGTGTAGGTGTCCGCCCTCGGCC-----AGACCCGCCAC | 14747 |
| Qy | 1773 | -----SerSerSerProLeuSerProGly----- | 1780 |

| | | | | | | | |
|----|-------|--|-------|----|-------|--|-------|
| Db | 18608 | GAGCTGGGGTTGTCGGAAACCGCAGCAGCATCCAGGTTCCGGCTCGGTGTCGAGGAG | 18549 | Db | 17536 | TCCTCCAGGACCG | 17501 |
| Qy | 589 | -----GluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaG1 | 603 | Qy | 839 | uGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaG1 | 859 |
| Db | 18548 | CACGAACCGGTCCGGCACTCGGCTCGCACAGCGGATCACGCCCCACATCGCGCGGC | 18489 | Db | 17500 | CGCAGCGGAGATCAGACAGGACCGTGCCTCCGCGCACCAGCTCCGGCGCCCGCAGCAGTGT | 17441 |
| Qy | 603 | uLeuAlaSerMetGluLeuLeuAsnGluSerSerArgTrpThrGluGluMetGluThrAl | 623 | Qy | 859 | uGluProValLysSerGluCysThrGluGluAla | 873 |
| Db | 18488 | CGCGGGTTCGGAAACCTGCTCGTCCGACACGGTGACCGCGCGCGCGCGATGAC | 18429 | Db | 17440 | TCGGCAGCCAGCGGTCTCGTCCGTGGAAAGCTTCCTCGGTGAACACCGCCACCGCGGTGG | 17381 |
| Qy | 623 | aLys-----LysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaAr | 639 | Qy | 873 | yProAlaLysGlyLysAspAlaGluAlaGluAlaThrAlaGluGlyAlaLeuLysAl | 893 |
| Db | 18428 | CAGGTGCGCGCCCGCAGCTCTCTGTTGAGCAGCAGCGCGCAGCGCGCGGCCA | 18369 | Db | 17380 | CCCGCAGCTCAGCCGCGCAGCAGCGGTCTCCGCGCCCGCAGCGCCCGCCAGCCG | 17321 |
| Qy | 639 | gMetValGlySerLysThrValSerGlnCysLysAsnPheTrpPheAsnTrpLysLysAr | 659 | Qy | 893 | agLysLysLysGluGlyGlySerGlyArgAlaThrAlaLysSerSerGlyAla----- | 911 |
| Db | 18368 | GAGGACGCGGTGGCGCGCGCGCGGGGTCTCGCTTCGGTTCGTAGCGGACGAC | 18309 | Db | 17320 | GAGACGTGCGCGGCGCGGT-----CGCGCCGCGCCAGAGCGGTGTTGTTGGAAC | 17270 |
| Qy | 659 | gGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAl | 679 | Qy | 912 | -----ProGlnAspSer----- | 915 |
| Db | 18308 | GACCGCTCGGTTCCGGTTACCGGACCGCAGCAGCGCGTGCAGGTCCGAGGCATCGC | 18249 | Db | 17269 | GGGTAGGTGGCAGCTCGCGCGCGCTCCCGCGGCACCGCCGCGACCGAGGTTCGGCGCG | 17210 |
| Qy | 679 | aArgArgLysLysLysAlaProAlaAlaAlaSerGluGluAlaAla | 695 | Qy | 916 | -----AspSerSerAlaThrCysSerAlaAspGlu-----ValAspGluAlaG1 | 930 |
| Db | 18248 | CGCGACCAACACCGCGTCCGAGCGCGGTCTCGGTAGCGTCCCGCAGTCCAGCGTGCAG | 18189 | Db | 17209 | ACGCCGTGCACGTGACCGCGCAGAGGTTCGCGAGGAACCTCGTCGCTCGCTCTGCTCG | 17150 |
| Qy | 696 | -----PheProProValValGlu----- | 701 | Qy | 930 | uGlyGlyAspLysAsnArg-----LeuLeuSerProArgPr | 942 |
| Db | 18188 | CTCGCCCTCGCACCGCGCGGTTCGGGTGCTCCACCTCTCTCCCGTGTCTCGGACGAC | 18129 | Db | 17149 | CGCGCAGCGTCGGCAGCGAGGTTCGCGCGCGCTTCGCGCGGTCTCTCTCGAC | 17090 |
| Qy | 702 | -AspGluGluMetGluAla----- | 707 | Qy | 942 | oSerLeuLeuThrPro-----ThrGlyAspProArgAlaAsnAla-----SerProGlnLysPr | 960 |
| Db | 18128 | CAGCGAGTCAGCGTGGCGACACAGGTGGCTCGCGGTCTCGGTGACCCCGAGCGCATCGC | 18069 | Db | 17089 | GCCATCAGCAGCACCGCGGTGCGCGCTCCTCGACGAAGGTGCGAAGCCCTCTCGACC | 17030 |
| Qy | 708 | -----SerGlyValSerGlyAsnGluGluMetValGluGluAla----- | 721 | Qy | 960 | oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaIleLeuProProIleGlnValThrLy | 980 |
| Db | 18068 | GTCCGCGCGCGGGGTCCGACCGACCGCAGCGAGTCCGCCGCGAGCGGTGCAGGT | 18009 | Db | 17029 | A-----CACCGCGCA-----CCGCTCTCGAAGACACT | 17000 |
| Qy | 722 | -GluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGlyCysSerGlyPro----- | 739 | Qy | 980 | sVal----- | 982 |
| Db | 18008 | CACGGTGTCCAGCGAAGCGCAGCTTCCCGCGCGCGCTCGCGAGCGCGCCAGGCT | 17949 | Db | 16999 | GGTCGGCGCAGGTTCGGGTACAGTAGGAGCGCTCCATCGCGAGGTGTGATGACTCG | 16940 |
| Qy | 740 | -----AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisTh | 756 | Qy | 982 | sGluProProArgGluAspAlaAlaProThrLys-----ProAlaPr | 996 |
| Db | 17948 | GAGCGTCTGGCCACCGCGTCCGAGCAGCAGCGGTGGAACGCTAACCTTCC-----TC | 17895 | Db | 16939 | CCGGTACCGTGGAGTGCAGCGGCACCGAGCCCGCGCAGCGGTGATGGTCCAGCTCG | 16880 |
| Qy | 756 | xGluAlaAlaLysAspThrGly----- | 763 | Qy | 996 | oPro-----AlaProProProProGlnAsnLeuGlnPro-----G1 | 1008 |
| Db | 17894 | GTCCGCGCGCATGACACCTCGCGGTAGACCGAGTCTCTTTCGCGCACCGGCTCGCAG | 17835 | Db | 16879 | GCCGCGAGCGGTCTCTCGATCGCGCCACCTCCGGGAGTGGAGGCGTAGCGCACGCGC | 16820 |
| Qy | 764 | -----GlnAsnGlyProLysProProAlaThrLeuGly----- | 774 | Qy | 1008 | uSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPr | 1028 |
| Db | 17834 | GGCTGGAACAGCGTCCGTACTCGTA-GCCCTGTCTCGCGAGCGCGTGTAGTGTCTGT | 17776 | Db | 16819 | ATCCGCGCGGTTCATCTCCCGCGCTCGCGCTTCGCGGAGAACTCGT-----CCAGTCC | 16763 |
| Qy | 775 | -----AlaAspGlyProProProGlyPro-----ProThrProPro-- | 786 | Qy | 1028 | oProAlaAspLysGluAla-----PheAlaAlaGluAlaGlnLysLeuPro-----G1 | 1044 |
| Db | 17775 | CGAGCGGATGCGGACGCGCT-CCTCCGCGCGCCACTGCTCGGTATCCCTCGGACCGCG | 17717 | Db | 16762 | CGGTGCGCGCGGCAACGACGCTGTCAGGCCCTTCACGGCGCGCACCGCACACGG | 16703 |
| Qy | 786 | ----- | 786 | Qy | 1044 | yAspProProCysTrpThr-----SerGlyLeuProPheProVa | 1057 |
| Db | 17716 | GCGAGCGCTCCGCGAGCGTCCCGTGGGTGCTGGACCACTCGCGCTCGCGAGGTCTG | 17657 | Db | 16702 | CGTCCAGCGCGGAGCAGCGTCTCCAGTCTGTCGGGAGCGCGCGCACCATG | 16643 |
| Qy | 787 | -----ArgArgThrSerArgAlaProIleGluProThrProAl | 799 | Qy | 1057 | lProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTy | 1077 |
| Db | 17656 | GCCACGCTTTGGCGCGGTGAGCTGATCGTCCGCCGCCCACTGCTCGCGCGCGCG | 17597 | Db | 16642 | CCGCCCTCGTCTGAGTCTCGCGCAACCTTGGTGGTGGCGAGGCCA----- | 16597 |
| Qy | 799 | asrGluAlaThrGlyAlaProProProProAlaProProSerProSerAlaProPr | 819 | Qy | 1077 | rAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuPr | 1097 |
| Db | 17596 | ACCACATGCGCAGCG | 17537 | Db | 16596 | -----CCACTTCGCGCGCTTCCAGCG-----ACAGCGCGCGCG-----CC | 16559 |
| Qy | 819 | oProValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluG1 | 839 | Qy | 1097 | oArgProProThrIleSerAsnPro-----ProProLeuIleSe | 1110 |
| | | | | Qy | 16558 | ACGTGCGCGCGCGATCTCGCCCTCGCGAGTGGCGCGCGCGCGCGCGG-GTACGCC | 16500 |

```
QY 13 ArgAlaThrGluProArgTyrProHisSerLeuSerTyrProValGlnIleAla--- 31
Db 20514 CGCGGAGCGTACCGCGAGACCGCGCGGTGTGAACGACACAGTTCGCGCCAGCCGG 20455
QY 32 -----ArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp 48
Db 20454 CGACCTGTGTCGAGTCCGACCGGACC--GGC-----GAGCGGTTCGACACAGTTCGGCCC 20403
QY 49 TyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeu 68
Db 20402 TGGCGCTCGGCTTCGCGGCGCGCGACGCTTCGCTTCGACCATGCGCGCGAGGACCTC 20343
QY 69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSer 88
Db 20342 CGGCACGAATCGCGCCCTCGCGAGCGCGGAACGGTTGAT----- 20304
QY 89 HisSerTyrLeuProGluLeuGlySerGluMetGluPheIleGluSerLysArgPro 108
Db 20303 -----GGACAGCGGAACACACACCTCACCGCGCGCGCGAGGC 20265
QY 109 ArgLeuGluLeuLeuProAspProLeuLeuArg-----ProSerProLeuLeu 124
Db 20264 GCTGTGAAACAGTCGAGCGCCGCTCGTTCGCGAGCGCGCGACCCCGGTTCGCGCGAT 20205
QY 125 AlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys 144
Db 20204 GCGGTGCGCGAGTCCGCTGT-----CATCTCGCTGGC--- 20172
QY 145 LeuGluProValSerProProSer-----ProProHisThr 156
Db 20171 CTGCGCCACAGTCCCATCCGAGCGCTTCGCGGGAAGCGCGCGCTCTCCGCGAGGC 20112
QY 157 AspProGluLeuGlu---LeuValProProArgLeuSer----- 168
Db 20111 AGCCAGCGGTGAGCGACTCGTTTCGCGCGCGGTACACGCCCTGCCCCGGCGCGGTAG 20052
QY 169 -----LysGluGluLeu 172
Db 20051 CACCGACCGCGGAGAGAACAGCAGCAGAGAGCTCAGACCGGTGTTCGCGGTACGCTC 19992
QY 173 IleGlnAsnMetAspArgValAsp-----ArgGluIle 183
Db 19991 GTGCAGGTTCACGCGCGGTGCGACCTTGGCGCGCAGCACCTGTCTCCACCGCGCGCTCGTC 19932
QY 184 ThrMetValGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGlu 203
Db 19931 GATGGAGGTGACGACCGCGTCCGC-----CAGCACACCGCGCGGTGCGACAGACC 19881
QY 204 AlaAlaLysProGluPro-----GluLysProValSerProProIle 219
Db 19880 GGTACGCGCGCGGCGAGCCGCTCCAGCAGCGCGGAAGCGGTTCGCGGTTCGCGGTGTC 19821
QY 220 GluSerLysHis---ArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAla 238
Db 19820 GCAAGCGCGATCTCCGCGAGCGCGCCAAAGTCTC-----GATCTCCGCGCGCGAGCTC 19767
QY 239 GluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro----- 255
Db 19766 GTCGGAACCGCGCGGTTCGCGCGCGCGCTGTGACGAGCAGAGGTGCGCGCGACCGCGGTG 19707
QY 256 -----LeuTyrAsnGlnProSerAsp-----ThrArgGlnTyr 266
Db 19706 CTCGGTGACGAGTGGCGCGAGCAGCCGTCGAGCGGTGCGCGGTGCGCACCGGTATCAG 19647
QY 267 HisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArg 286
Db 19646 CACGTCGCTCCGCGT----- 19629
QY 287 ArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGlnLeuMet 306
Db 19628 -----CACCGCGCGCGGTGGGTTCAGCACAGCTTCGCGAC----- 19593

QY 307 GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu 326
Db 19592 -----GTGCTCTCCCGGCTCATGTGTCTGCAACGCGCGCGCGGCGGA 19551
QY 327 SerLysVal---ArgGluTyrTyrGluLys---GlnPheProGluIleArgLys----- 342
Db 19550 TCCAGCTCTCCAGCGCATACCGGAGCGGTTCGAGCTCCCGCGCGCCAGAGGCGGAC 19491
QY 343 -----GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db 19490 GACCTCGCGCAGGATCTCCCGAGCGGTCTGTCACCGCTCTCCCGAGGTTCGACGCGGC 19431
QY 361 LeuSer-----MetSerAlaAlaArgSerGlu 369
Db 19430 GTATCGCGCCCGGAAGTCCCGCGCTCCCGCAGGTCTTGTCCCATCTTCGACGAAGC 19371
QY 370 HisGluValSerGluIleIleAspGlyLeuSerGluGln-----GluAsnLeuGlu 386
Db 19370 GCGTCTCTCGCGAGCAGATCCCGGACTCTGTGAGCAGTTCCTCCGCTGAGCGAGTTGAG 19311
QY 387 LysGlnMetArgGlnLeuAla-----ValIleProMetLeuTyrAspAlaAspGln 404
Db 19310 CACCACTCCAGCGCGCGCTCCGGTGGCTCCCGAACTTCGCGCGGAA----- 19260
QY 405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMet-----AlaAspProMetLysVal 422
Db 19259 -----ACCGGTCTCCCGGAGAGCGATGTCTGTCTCGTCTC 19224
QY 423 TyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPheArgGlu 442
Db 19223 GAGACCGAGCGC-----CCGACGCGTCCCGTGTTCGCGCGGCC 19185
QY 443 LysPheMetGlnHis-----ProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db 19184 GCGGTGGCGCAACACCTCCGCCCGCTCCGCGCGC-----CAGCGC 19143
QY 461 -----LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLys 475
Db 19142 GACGCGCGCATCGCGACACCGCTCGCTGCGTGTGATGAGCACCGACTGACCGCGCG 19083
QY 476 AsnGluAsnTyrLysSerLeuValArgSerTyrArgArg----- 489
Db 19082 CAGCCC-----CGCAGATCTGCGCGCGGTAAATGCGCGGTGTGAGCGATGGGCAC 19029
QY 490 -----ArgGlyLysSerGlnGln----- 496
Db 19028 GCGCGCGCGTTCGCGCTCCAGCGCTCCGACACTCGTTCGCGAGGCGCGTATCGGT 18969
QY 497 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSer 516
Db 18968 GACCGGATCGCGCGAAGGCTCCCTGGAACAGCGCGAGCACCGGTCTCCCGCGCGGAA 18909
QY 517 -----GlnGluGluLysAspGlu-----Lys 523
Db 18908 GCGTCCAGCTCCGCTCCGACCGCGCTGAGCAGCACCGCGCGCTCGGTGCGCATGTCGC 18849
QY 524 GluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAsp----- 541
Db 18848 CTTCTCGGATACATCCGAGTCCGAGGAGGAGCTCGCGAAGTTGACTCCGTTGGCGCG 18789
QY 542 -----LysGluAspLeuLeuLysGluLysThr-AspAspThrSerGlyGluAspAsnAs 559
Db 18788 CAGTCCACCGCGACCTCCGCGCGCGCGGCGGTCTGTCACCTCGGTTCGCGGTTCGAA 18729
QY 559 pGluLysGluAlaValAla---SerLysGlyArgLys----- 570
Db 18728 GGCACGAGTTCATCGCGCGCTCACCCCGCAGCAGACGTTGGGTGCTCTCCCGAAGCGT 18669
QY 571 -----ThrAlaAsnSerGlnGlyArgLysGlyArgIleThrArgSerMetAlaAsn-- 588
Db 18668 CAGCGGGAAGGTTCGAGCGCGCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 18609
QY 588 ----- 588
```

2205 -----ArgSerProGluProAsnLysT 2212
21413 GGCCACCGATCCGCGCGGACGCGCGGAGTCCGCGACCGCACTGGCGGAA 21354
2212 hrSerValLeuGlyGlyGlyLeuAspGlyLeuProValSerProGluGlyMetT 2232
21353 CTCGAACCGCAGCAGCA-----GCACGGGACACACGCGCCACACGCGCGAGGT 21303
2232 hrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnT 2252
21302 GAAGCCCATGTCGAGT-----A 21285
2252 hrGluProSerArg-MetGlySerLysSerProGlyAsnThrSerGlnPro---ProAla 2270
21284 GCGGCGCAGCGCGAGTTCAGTCAGCGTGTCAATCGACGTCCTTCGCGTCCCTCGC 21225
2271 PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluLe 2290
21224 CCGCTTCGCCATGTCGGGAC-----CGCGCTCAAAACCGCCAGC 21180
2291 AsnLysLysLeuAsnThrHisAsn---ArgAsnGluProGluTyrAsnIleSer---Gln 2308
21179 AACCGCCGCTCAGCGCTCGAATCTCGCGGCCCGCCAGGAGTTCGCGTCCGCGTCAG 21120
2309 ProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArg 2328
21119 CCGCCAGCTCGAGCTTTACCGTACCGCGGTCTCGGC-----GTGATCCGC 21072
2329 SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLys 2348
21071 GCCACACCGCGGCTCGCGACCGACGACGACGACGATCCCGCATCATCGTGATT 21012
2349 AlaLeuMetGlyLysTyrAsp-----GlnTrpGluGluSerProProLeuSerAla 2365
21011 GCGAAGACCGACCGCCAGCTGGCGCGGCTGCTGGAGACCTCGCAGGTTGACGACGCG 20952
2366 AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPro----- 2382
20951 ACCGCG-----CCGTGAACCGGACC---ACCGTCCGCGCGCGCGGCTTGTGTCGATG 20901
2383 -----IleThr-AlaAlaAspGlyArg 2389
20900 CGGACTTGTGACCGCGCGAGGTGACCGCTTTCTTTCGACGCGCGGCTTGTCTCG 20841
2389 gSerAspHisThrLeuThrSerProGlyGly-----GlyLysAl 2403
20840 ATCA-----GCCGGTGTCCAGGTGAACAGCACTGGCGCGC 20802
2403 alyValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGly--LeuAla 2422
20801 CCGCTGTCCAGCGCGTGTGCGCGCGGTAGATGTCGATGGCCACCGCGCGGCGATTGAGC 20742
2423 SerGly-----AspArgProProSerValSerValHisSerGlu----- 2436
20741 GCGGGGAAACGCGACGACGACGAGGTGCGCTGAGTCTCGTGGGAGCCAGCAGGCGCC 20682
2437 -----GlyAsp 2438
20681 TGGATCGCATCTGGTGTGGACCGCTCGTGGGTTGGGTAGCTCCGCGCGCGGG 20622
2439 CysAsnArgArgThrProLeuThrAsnArgValTyr-GluAspArgProSerSerAlaG 2458
20621 TCATGGCAGCAGCGCGCGGACAGAGGTCTGGCGGTTCGCGGTTCGCGCTGCACG 20562
2458 ySerThrProPheProTyrAsnProLeuLeuMetArgLeuGlnAlaGlyValMetAlaSe 2478
20561 GTGACGTCGCGGTCTCGCGTCCGGAACGTCACGCTGAAGTGGGCG-----CATAG 20508
2478 rProProProProGlyLeuProAlaGlySerGlyProLeuAlaGly 2493
20507 CCGTGGCCCTGCGAGGTACACCGGTTCGCGCGCGCGCGCGCGGT 20462

AAQ46806/c
ID AAQ46806 standard; DNA; 29879 BP.
XX
AC AAQ46806;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 22-DEC-1993 (first entry)
XX
DE eryA region of S. erythraea chromosome.
XX
KW Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW erythromycin; condensation; elongation; acyl chain growth;
KW gene replacement; ss.
XX
OS Saccharopolyspora erythraea.
XX
FH Key Location/Qualifiers
CDS 3..10199
FT /*tag= a
FT /note= "ORF 1"
CDS 10218..20921
FT /*tag= b
FT /note= "ORF 2"
CDS 20922..29879
FT /*tag= c
FT /note= "ORF 3"
XX
PN W09313663-Al.
XX
PD 22-JUL-1993.
XX
PF 17-JAN-1992; 92WO-US000427.
XX
PR 17-JAN-1992; 92WO-US000427.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Katz L, Donadio S, Mcalpine JB;
XX
DR WPI; 1993-242804/30.
XX
DR P-PSDB; AAR44430, AAR44431, AAR44432.
XX
PT Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -
XX by introducing altered biosynthetic gene-contg. DNA into microorganisms.
XX
PS Claim 27; Fig 2; 133pp; English.
XX
CC This sequence represents a fragment of the Saccharopolyspora erythraea
CC genome, designated eryA. The polypeptides encoded by this region are
CC involved in the biosynthesis of the polyketide segment of erythromycin.
CC eryA is organised in modules and each module takes care of one
CC condensation step. The precise succession of elongation steps is dictated
CC by the genetic order of the modules. This fragment may be specifically
CC altered such that novel polyketide molecules of desired structure are
CC produced. Three types of alteration may be produced; those inactivating a
CC single function in a module which does not arrest acyl chain growth;
CC those inactivating a single function in a module which does affect chain
CC growth; and those affecting an entire module. The mutations may be
CC introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN
CC field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.89e-09 Length: 29879
Score: 596.00 Matches: 710
Percent Similarity: 32.80% Conservative: 334
Best Local Similarity: 22.31% Mismatches: 1135
Query Match: 4.51% Indels: 1019
DB: 2 Gaps: 157

US-09-522-753-5 (1-2517) x AAQ46806 (1-29879)

Db 23417 GTTCGGTAGCGCTTCGCCGGTCTCCCGCGC-----CAACCGAT 23379
Qy 1685 eileAsnAspTyr-----ileThrSerGlnGlnMethis---HisAsnTh 1699
Db 23378 CCAGCGGATTTCCGGCGCTTCTTCGAGGTGTGGTCGAACGGCACCGCGCACCGCAC 23319
Qy 1699 rAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSe 1719
Db 23318 GGCAACGCGCT-----GGAGAAGTGTCTGACGACCTTGCC-----CACGTGCA 23277
Qy 1719 rLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleileAspLeuSerGlnValProHi 1739
Db 23276 CCTTGCC-----CGCGCGCGCCGACCGCGTTGACGACGCGGATCGAGTTGGCA 23226
Qy 1739 sLeuProValLeu-----ValProProThrProGlyThrProAlaThrAlaMetAs 1756
Db 23225 CCTGGTGTGTGGCGCGTGGCGCTTCGACCCAGT--CCAGCGTCCGCGCGCGCTGC--GG 23170
Qy 1756 pArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerPr 1776
Db 23169 GCGCGGTGTGATGGACCCGCGTGTACCGACGACGACGACGACGACGACGACGACG 23110
Qy 1776 oLeuSerProGlyGly----- 1781
Db 23109 GACATCGCCGAGTGCACAAACCCCGTACCGGATCGGTGAGCACCAGCGCGTGAACCGC 23050
Qy 1782 -----ProTh 1783
Db 23049 TCCAGATGTCACGGGTGCGGGTGTAGTCTCGCGCGTGAATCGTCCGAGGAGCGCGGA 22990
Qy 1783 rHisLeuThrLysProThr-----ThrThrSerSerSerGluArgGluAr 1798
Db 22989 TCACTGTCCGATGCCAGCACTTCCAGCCCGTGGTCCCTTCGATGTGTAGCGCGAAGC 22930
Qy 1798 q-----AspArgAspArgGluArgAspArgAspArgGluArg----- 1810
Db 22929 TCGGTGTAGCGCTCGATGCGCTCGCGCTCG-CCGACGAGGTGCGCGACCGTGGCGCGGT 22871
Qy 1811 -----GluLysSerIleLeuThrSerThrThrVa 1821
Db 22870 CTCGAGCGGACCTTTTGGCTCAGTGTGCGCGTGGGATGGCGTCCGACCGCAAC 22811
Qy 1821 lGlu-His-----AlaProIleTrpArgProGly----- 1830
Db 22810 CCGGTACACCGCAGGTGCGCAAGGTGCGCCCGCCAGGGTCCGCGTGTGCTTGTGCC 22751
Qy 1831 -----ThrGlnSerSerGlySerSerGlySerSerGlyGly----- 1843
Db 22750 GACACCCCTTTACACACACACCGCGATCCGCTAGCAGCCGCGCGCGCGGTGAC 22691
Qy 1844 -----GlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1861
Db 22690 CTGAGCGACCGAGTCTGTAGGTGCGCGCGCGAGGCGCTTCCCGCGGTGTATCAGCCG 22631
Qy 1861 erProIleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHisAsnT 1881
Db 22630 ACCCGACAGCAGATAGACCAACCGACGCC-----AGCCTTCGACGCGTGGCGGCTTG-A 22578
Qy 1881 hrGlyMetLysGlyIleileThrAla-ValGluProSerLysProThrValLeuArgSer 1900
Db 22577 CCGCGCTCAGCAGCGTGTGGTGGCCAGTTCCG-----GAT 22542
Qy 1901 ThrSerThrSerProValArg---ProAlaAlaThrPheProProAlaThrHisCys 1919
Db 22541 ACAGCGATGGGCGACCGCTCGAGACGATACGCGCGCTTTCAGCGCGGC----- 22491
Qy 1920 ProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuPro 1939
Db 22490 -----TGGCGC----- 22485
Qy 1940 LysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPhe 1959
Db 22484 -----GGCCCTTCATCGGGCGCGCTGTAGTAGGCGCGGTAGGCGGTGGGTTCGAGCC--- 22434

Qy 1960 LeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySer 1979
Db 22433 -----GCTGGGTGCGGAATCCCGCACTGACCTAGTCGGTGTTCACCCAC 22389
Qy 1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThr----- 1997
Db 22388 CGCCCGCAGCGCGTGTSCCCCCCGATGGCAGCGGCGCAACTGCACCGTCCGATGAT 22329
Qy 1998 ProAlaLysAsnLeuAlaProHisHisAlaSerProAspPro----- 2011
Db 22328 CCGTGGCGCGT---TGCCGCTATCGGGGCTCCACCGCCGCTCCTCGGATGTCGGTCA 22272
Qy 2012 ProAlaProProAlaSerAla-----SerAspProHis 2022
Db 22271 CCAGCACTCCCGCTGTGCGGCTCTCGGTGCGTTCGACGACGAAACGGTCAGCCACGCC 22212
Qy 2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly 2042
Db 22211 CGCATAGATGACGCGCGCGACGCTCACAGGTTTAACGAAGAACAGATTGAAGACCATGGC 22152
Qy 2043 TyrHisGlySerSerTyrSerProGluGly-----ValGluProVal----- 2056
Db 22151 CACGAACCGCGCACATTGATCCAGCGGCTTCGGTCCCGCAACCCCGCTTGACCG 22092
Qy 2056 ----- 2056
Db 22091 CGCGTGCAGGTACCGCGGTACACCCACCGACGAGGACACCGCTCTCTTTGGGGTC 22032
Qy 2057 ---SerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlu 2075
Db 22031 CCAGCCCCAGTAGCGCGCGCGCGCTTCTCGCGCCAGATGCCACCGAATATCACCCCGAA 21972
Qy 2076 GluLeuAspLysSerHisLeuGluGlyLeuLeuArgProLysGlnProGlyProVal--- 2094
Db 21971 -----GCCAAACCGGGGAGCGAAGATCGT-----GTCGCTAGGC 21933
Qy 2095 -----LysLeuGly---GlyGluAlaAlaHis-LeuProHisLeuArgProLe 2109
Db 21932 GATTCGCTCCAGGTTTGGCATCGGAGCGCGCACCATACCCCGCAGCGCGCTTC 21873
Qy 2109 uProGluSerGlnProSer-----SerSerProLe 2119
Db 21872 ACCGGTGTGCTACCCAGCGCGATGTGCGCACAAAGAACAGGATGTGCGCACCGCC 21813
Qy 2119 uLeuGlnThrAlaPro----- 2124
Db 21812 GACGAGATATCCCGAAGCCAGGCTGACCCGACACATGAATGGGACCGCATAGGA 21753
Qy 2125 -----GlyValLysGly-----HisGlnArgValValThrLeuAl 2136
Db 21752 CTGAGTGCAGCGCATCACCGGGCGCATTTGGGTAGAGCAGCGTCCGACACG----- 21698
Qy 2136 aGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisPro----- 2152
Db 21697 -----GTGACGAGATCAGCAGCGGACACGACGAGAACCCACACGCGCGCGTATCG 21645
Qy 2153 -----GlnGlnLeuSerAlaProLeuProAlaPro----- 2162
Db 21644 CGACCGCGGACGACACCGCGCGGATGACCCCGGACAGCGATCAGGTAGTTGATGAA 21585
Qy 2163 -LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProSerAs 2182
Db 21584 CTCGTATATGTTCCCGACCGCGCGCTGGGTG-----GCCAGCGCGCGCAGCAC 21534
Qy 2182 pLeuTyrLeu-----ProPro-ProAspHisGlyAlaProAlaArg----- 2195
Db 21533 GACGCGCAGCAGTAGTACCGCGATGCGCGACGCGCGCGCGCGCGCGCGCGCGAC 21474
Qy 2196 -----GlySerProHisSerGluGlyLys----- 2204
Db 21473 GCGTTCGTGACCGCGCGGTGCGTGTGTCGCGCACGATCCAGGGGTAGCGGTATCGGT 21414

[illegible]

Db 33451 CCAGGCCCGGCGATCTCCGGG-----CGCGCGGGAGCGCCCGCTGCTCTGCTCGG 33504
 Qy 2437 ly-----AspCysAsnArg-----ThrProLeuThrAsnArgValT 2450
 Db 33505 GCACTCGGCGGCGCCCTGCTCGCGACGAGCTGGCTTCCGCTGGAGCGGCGCACGG 33564
 Qy 2450 rpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetA 2470
 Db 33565 CGCGCCCGCGCGGATCGTCTGTCGACCCCT-----ATCCGCC 33606
 Qy 2470 rgLeuGlnAlaGlyValMetAlaSerProProProGlyLeu-ProAlaGlySerGly 2489
 Db 33607 GG-----GCCATCAGAGGCCCATCGAGGTGTGAGCGAGCGAGTGGCGGAGG 33654
 Qy 2490 ProLeuAlaGly 2493
 Db 33655 CCGTTCGCGGG 33666

RESULT 90

AAI99683_06/c
 Continuation (7 of 44) of AAI99683 from base 600001 (Mycobacterium tuberculosis strain H
 WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

| Fragment Name | Begin | End |
|---------------|---------|---------|
| AAI99683_00 | 1 | 110000 |
| AAI99683_01 | 100001 | 210000 |
| AAI99683_02 | 200001 | 310000 |
| AAI99683_03 | 300001 | 410000 |
| AAI99683_04 | 400001 | 510000 |
| AAI99683_05 | 500001 | 610000 |
| AAI99683_06 | 600001 | 710000 |
| AAI99683_07 | 700001 | 810000 |
| AAI99683_08 | 800001 | 910000 |
| AAI99683_09 | 900001 | 1010000 |
| AAI99683_10 | 1000001 | 1110000 |
| AAI99683_11 | 1100001 | 1210000 |
| AAI99683_12 | 1200001 | 1310000 |
| AAI99683_13 | 1300001 | 1410000 |
| AAI99683_14 | 1400001 | 1510000 |
| AAI99683_15 | 1500001 | 1610000 |
| AAI99683_16 | 1600001 | 1710000 |
| AAI99683_17 | 1700001 | 1810000 |
| AAI99683_18 | 1800001 | 1910000 |
| AAI99683_19 | 1900001 | 2010000 |
| AAI99683_20 | 2000001 | 2110000 |
| AAI99683_21 | 2100001 | 2210000 |
| AAI99683_22 | 2200001 | 2310000 |
| AAI99683_23 | 2300001 | 2410000 |
| AAI99683_24 | 2400001 | 2510000 |
| AAI99683_25 | 2500001 | 2610000 |
| AAI99683_26 | 2600001 | 2710000 |
| AAI99683_27 | 2700001 | 2810000 |
| AAI99683_28 | 2800001 | 2910000 |
| AAI99683_29 | 2900001 | 3010000 |
| AAI99683_30 | 3000001 | 3110000 |
| AAI99683_31 | 3100001 | 3210000 |
| AAI99683_32 | 3200001 | 3310000 |
| AAI99683_33 | 3300001 | 3410000 |
| AAI99683_34 | 3400001 | 3510000 |
| AAI99683_35 | 3500001 | 3610000 |
| AAI99683_36 | 3600001 | 3710000 |
| AAI99683_37 | 3700001 | 3810000 |
| AAI99683_38 | 3800001 | 3910000 |
| AAI99683_39 | 3900001 | 4010000 |
| AAI99683_40 | 4000001 | 4110000 |
| AAI99683_41 | 4100001 | 4210000 |
| AAI99683_42 | 4200001 | 4310000 |
| AAI99683_43 | 4300001 | 4403765 |

Alignment Scores:

Pred. No.: 4.47e-09 Length: 110000
 Score: 598.00 Matches: 503
 Percent Similarity: 32.81% Conservative: 207
 Best Local Similarity: 23.24% Mismatches: 772

| Query Match: | 4.53% | Indels: | 691 |
|--|-------|---------|-----|
| DB: | 4 | Gaps: | 102 |
| US-09-522-753-5 (1-2517) x AAI99683_06 (1-110000) | | | |
| Qy 681 AtcLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700 | | | |
| Db 25925 CGAAGGTTGAGCGCGCGAAGATCGCGCTAGCAGGATCAGCCGTG-----25878 | | | |
| Qy 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720 | | | |
| Db 25877 -----CGGCGGCTGTCGCGGGCTCCGACAG-----25848 | | | |
| Qy 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740 | | | |
| Db 25847 -----CGAACCGCGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 25803 | | | |
| Qy 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760 | | | |
| Db 25802 GGAAGCGCCAAACCGG 25743 | | | |
| Qy 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGly-----777 | | | |
| Db 25742 AGCAGCTGGGCATCACCGG 25683 | | | |
| Qy 778 -----ProProGlyProProProThrProProArgArgThrSerArgAla 792 | | | |
| Db 25682 CGCGCGCGGCACACCGCGCGCGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 25623 | | | |
| Qy 793 ProIleGluProThrProAlaSer-----GluAlaThrGlyAlaPro 806 | | | |
| Db 25622 CCCGACACCGG 25563 | | | |
| Qy 807 ThrProProAlaProProSerProSerAlaPro-----ProProVal 821 | | | |
| Db 25562 CGG 25503 | | | |
| Qy 822 ValProLysGluLysGluGluThrAlaAlaAlaProProValGluGluGlu 841 | | | |
| Db 25502 CCACCG-----ACAGCCCATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 25464 | | | |
| Qy 842 GluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluPro 861 | | | |
| Db 25463 CGAAGCGGCACCGG 25405 | | | |
| Qy 862 ValLysSerGluCysThrGluGluAlaGluGluGlyPro-----AlaLysGly 877 | | | |
| Db 25404 -----CCCACACCGG 25354 | | | |
| Qy 878 LysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGlu 897 | | | |
| Db 25353 GAAGACAGACCGCGCGGCACCGG 25294 | | | |
| Qy 898 GlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSer 917 | | | |
| Db 25293 -----GCCCGCACCGG 25249 | | | |
| Qy 918 Ser-AlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLe 937 | | | |
| Db 25248 AGCATCCACT-----25239 | | | |
| Qy 937 uLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPr 957 | | | |
| Db 25238 -----CGCGCTCCGG 25186 | | | |
| Qy 957 oGlnLysProLeuAsp-----LeuLysGlnLeuLysGlnArgAlaAlaAlaIleProPr 975 | | | |
| Db 25185 ACCATCACCACCGG 25126 | | | |
| Qy 975 oIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAl 995 | | | |
| Db 25125 GGCCCCCCCCATTACCAAGCAACCAACCA-----GCACCGCGCGCGCGCGCGCGCGCGCGCGG 25075 | | | |

Db 29442 A-CTCGCTGACGGCGGTGACAGCTCCGCAACAGCTCTCCACGGTGTGTCGGCAA-----C 29494
Qy 1288 InCysSerLysGluAspGly-----ArgSerSerSerGlyProP 1301
Db 29495 AGCTCCCGCCACACCGGTCTTCGACACCCGACGCGCCGCGCACTCGCGCGCACCTC 29554
Qy 1301 roHis-----GluThr-AlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
Db 29555 CACGAGCGTACCTCGCACCGCGCGACCGCGCCGCGACG-----GACTGGGAGGGCGG 29608
Qy 1318 ValGlyArgAlaIleSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db 29609 GTGGCGGGCCCTGGCCGAACCTCCCTCGACCGGTGCG-AGACGCGGGGTCTCTCGA 29667
Qy 1338 GluArgHisSerProHisLeuLysGluGlnHisIleArg-GlySerIleThrG1 1357
Db 29668 CACCGTCTCGGCTCACCGCATCGAGCCGAGCGGGGTTCGGCGGTTCGAGCGCGG 29727
Qy 1357 nGlyIle---ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLys 1376
Db 29728 CGCGCGCGACCTGTGTGCGAGCGGAGCGGTGATCGACGACCTGAGACCGCGAGGCCCT 29787
Qy 1376 sLeuLeuLysArgGluGly----- 1382
Db 29788 GATCCGGATGGCTCTCGGCCCCCGTAACTGACCGCGCGGTCTCTCCCCACGCGC 29847
Qy 1383 -----ThrProProProProProSe 1390
Db 29848 CGCACCCCGCATCCCGCGCACCAACCCGCGCCACACGCGCCACACCCCATCCAGCAGC 29907
Qy 1390 rArgAspLeuThrGluAlaTyrLysThr-----GlnAlaLeu-GlyProL 1405
Db 29908 GGAAGACACACACGATGACGAGTTCACACGACACAGTTCGTGAGCGCTCTGCGCGCCTC 29967
Qy 1405 euLysLeuLysProAlaHisGluGlyLeuValAlaThrVal-----LysG 1420
Db 29968 TCTCAGGAGAACAGAACTCCGGAAGAGCGCTCGCGCGCGCACCGCTCGGCGAGGA 30027
Qy 1420 luAlaGlyArgSerIle-HisGluIleProArgGluGluLeuArgHisThrProGluLeu 1439
Db 30028 GCCATGGCGATCGTCGGCATGAGTCCCGTTCGCGGGCGGAATCCGGTCCCGCAGGA 30087
Qy 1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr----- 1455
Db 30088 CCTCTGGAGCGGTGCGCGCGGCAAGGACCTGGTCTCCGAGGTACCGGAGGAGCGCGG 30147
Qy 1456 -----ProLeuLysTyrAspThrGlyAlaSerThrGlySerLysLys 1470
Db 30148 CTGGGACATCGATCCCTCTACGA-----CCCGGTGCGCGCGCGCAAGGG----- 30192
Qy 1471 HisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeu 1490
Db 30193 CACGACGTACGT-----CCGCAACGCGCGCTTCTCTCG-ACGACGCGCGCG 30236
Qy 1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer 1510
Db 30237 GATTTCAGCGGGCTTCTTCGGGATCTCGCGCGGCGAGGCCCTCGCATGGAGCCCGCAGC 30296
Qy 1511 Arg-----ProGlyThrAlaSerSerSerGlyGlySerIleAlaArg 1524
Db 30297 AGCGGCAGCTCTCTCGAAGCTCTCTGGAGGTCTTCGAGCGGGCGCGCATCGACCCCGCT 30356
Qy 1524 ----- 1524
Db 30357 CGGTCCGCGGCACCGAGCTCGCGGTGTACGTGGGTGTGGCTACGAGGACTACGCGCGCG 30416
Qy 1525 -----GlyAlaProValIleValProGluLeuGlyLysPro----- 1536
Db 30417 ACATCCGGGTGCGCCCGAAGGACCGCGGTTCAGTCTGTCACCGGCAACTCTCTCGCGCG 30476
Qy 1537 -----ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAla 1551
Db 30477 TGGCCTCCGGGGCATCGCGTACTCTCTCGGCTTGAGAGGAGCCCGCGGTGAGACA 30536

Qy 1552 GlyHisLeuPro---ArgGlySerProValThr----- 1561
Db 30537 CGGCGTCTCTCTCTCGTCTGCTGCCTGCACCTCGCCTGAAGGCGCTCGGAACGGCG 30596
Qy 1562 -----MetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
Db 30597 ACTGCTCGACGGCACTCGTGGCGCGTGGCGTCTCTCGCGACGCGCGGCGGTCTCA--- 30653
Qy 1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
Db 30654 TCAGTTTCAGACGACGAGCCATGCGCCGACCGCGCGGACGAGG-----GCT 30704
Qy 1595 SerPro-----HisSerThrValProGluHisProHisProIleSer-ProTyrG1 1612
Db 30705 TCGCTCTCGCGCGGCGCGCTCGCTGGGCGAGGCGTGGCGTCTCTCTCTCAAC 30764
Qy 1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
Db 30765 GGCTCTCCGACGCGCGCGCAAGG----- 30789
Qy 1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuPr 1652
Db 30790 -----CCACCGGT---CTGCGCGTGTGCGCGGCGAGCCATCAACC 30830
Qy 1652 oArgHisLeuAlaProAsnPro-ThrTyrProHisLeuTyrProProTyrLeuIleArgG 1672
Db 30831 AGGACGCGCGAGCAACGGCTCACGGCTCCGAC----- 30865
Qy 1672 lyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrS 1692
Db 30866 --GGGCGCTCCACGACGCGCTG---ATCGCGCAGCCCTTGGCGGCGCGCGGTCACT 30920
Qy 1692 erGlnGlnMetHis-----HisAsnThrAlaThr----- 1701
Db 30921 CGAGCCAGCTGGAGCTGTGGAGGCGCACGCGGAGCCCTCTCTCGGCGACCCGATCG 30980
Qy 1702 --AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro-----ArgG 1717
Db 30981 AGGCGCAGCGCTGTCTCGCACCGTACGGGCGAGGCGCGCGCGGCGCGCGTGGCG 31040
Qy 1717 luSerSerLeuAlaLeuAsnTyr-----AlaAlaGlyProArgGlyIleI 1732
Db 31041 TGGGAGCGCTGAAGTGCMACTCGGCGCACGCGAGCGCGCTTCGGGTGTCCGCGGTCA 31100
Qy 1732 leAspLeuSerGlnValProHisLeuProValLeuValProPro-ThrProGlyThrPro 1751
Db 31101 TCAAGATGTGCGAGCGCTCGCGCACCGGGTGTCTG-----CCGAGACCTCGCACGTGG 31154
Qy 1752 AlaThrAlaMetAspArgLeuAlaTyrLeuProThr-----AlaProGlnPro 1767
Db 31155 ACGACCGCGACGACCGAGTCTGCTGTGGCGCGGTTCGGTTCGAGTCTCACCGAGGCGC 31214
Qy 1768 PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLys 1787
Db 31215 TGGACTGGCGCGGA--CGCGCGCGCGCGCTCCCGCGGCGCGGTCTCCCGGTTCGGCGT 31272
Qy 1788 ProThrThrThrSerSerSerGluArgGluArg----- 1798
Db 31273 -----GGGCGGAGACGACGCGGACGCTCTCTTGAGGAGGCGCGCGCGGTCTGA 31320
Qy 1798 ----- 1798
Db 31321 GGAGTCCCTCGCTCGAGCGCGCGGTGGTGGCGGTGGTGGCGGTGGTGGTCCGC 31380
Qy 1799 -----AspArgAspArgGluArgAspArgAspArg-G1 1809
Db 31381 GAAGACCTCGCGCGCACTGGACCCCGAGTCTCGGCGAGTCTCGCGCATACCGCGAAGACCG 31440
Qy 1809 uArgGlu-----LysSerIleLeuThrSerThrThrThrValG1 1822
Db 31441 CACGAGCTGGATCCGGCGGTGGCGCGCGCTGGTTCGACAGCGCTACGCGCGATGA 31500

Db 25462 CTTCCGGGATCTCGCGCGCGAGGCCCTCGCCATGGACCCCGACAGCAGTGTCTCTCAC 25521
QY 209 -----GluProGluLysProValSerProPro 217
Db 25522 CACCGCGTGGAGCGCATCAGAGCGCGGCATCGACCCCGCGCCTTGAAGGCGCAGCG 25581
QY 218 ProIleGluSerLysHisArgSerLeuValGlnIleTyrAspGluAenArgLysLys 237
Db 25582 CTTCCGGCGTCTTCGTCGGCGGTGGCACACCGGTACACCTCGCGGCGAGCACCGCGCT 25641
QY 238 AlaGluAla-----AlaHisArgIleLeuGluGlyLeu-----Gly 249
Db 25642 GCAGTCCCGCGAGCTGAGGGCCACCTGGTCAGCGCGCGCGCTGTCTCTCTCTCGG 25701
QY 250 ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAen 269
Db 25702 CCGTATCGGTACGTCTCTCGTACGGAGCGACCGCCCTGACCGTGGACACGGCCTGCTC 25761
QY 270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289
Db 25762 GTCTCGCTGGTGGCTGCACCTCGCGTGCAGGCCCTCGCAAGGGCGAGTGGACAT 25821
QY 290 AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309
Db 25822 GGCCTCGCGGTGGTGTACGCTCATGCCCAACG----- 25857
QY 310 GluLysLysValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysVal 329
Db 25858 ---GGACCTGTTGTGCAGTTCAG-----CCGGCAGCGGGGTGGCCCGGACGG--- 25905
QY 330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg 349
Db 25906 -----CCGGTCAAGGGCTTCGCCACCTCGCGGACGG 25938
QY 350 MetGlnSerArgValGlyGlnArg-----GlySerGlyLeuSerMet-SerAlaAlaAr 367
Db 25939 CTTCCGCCCCCGAGGGCCCGGAGTCTGCTGGTGGAGCGCTGTTCGAGACCGCGCG 25998
QY 367 gserGluHisGluValSerGluIleIle-----AspGlyLeuSe 380
Db 25999 CAACGGACACCGGATCTCTCGCGTCTCGCGGCGAGCGCGGTCAACAGGACGGCGCCAG 26058
QY 380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTy 400
Db 26059 CAACGGCTCACGCTCCGACGCGGCCCTCCAGCAGCGGTTCATCCAGCGGCGCTGGC 26118
QY 400 rAspAla-----AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLe 415
Db 26119 GGACGCGCGCTCGCGCGCGGTGACGTGGAGCGTCTCGAGGCGCGACGGCAGCGCG 26178
QY 415 uMetAlaAspProMetLys-----ValTyrLysAsp-ArgGlnValM 429
Db 26179 GCTCGGCGACCCATCGAGCGCGAGGCCCTCATCGCCACCTACGGCCAGGAGAGCAG 26238
QY 429 etAsnMet-----TrpSerGluGlnGluLysGluThrPheArgGluLysPheMetG 446
Db 26239 CGAACACCGCTGAGGCTGGCGCGTGAAGT-----CGAACATCGGGCACAC 26286
QY 446 lnhisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluC 466
Db 26287 GCAGCGCGCGCGGTGTCCAGGTGTTCATCAAGATGGTCCAGGCGATGCCACCGGACT 26346
QY 466 ysValLeuTyrTyrTyrLeuThrLysLysAsnGluAenTyrLysSerLeuValArgArg 486
Db 26347 GCTGCG----- 26351
QY 486 erTyrArgArgArgLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db 26352 ----CGAAGACGCTGCATCGACGAGCCCTCGGACCATCATCGTGGTGGCGGCGAC 26406
QY 506 lnglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 526
Db 26407 GGTGGAACTCTCACCAGGCGCGTTCGACTGGCCG-GAGAAAGCAGGAGCGCGGCTCGCC 26465

QY 526 lu-----LysGluA 529
Db 26466 GC CGCGGTGTCTCTCTCGGCATCAGCGGGAGCAAGCGCAGCTGTCTCTGAGGAGG 26525
QY 529 laGluLysGluGluGluLysProGluValGlu----- 539
Db 26526 CCCCGCGGTTCGAGGACTCCCGCGCGCTCGAGCGCGCGGTGTGTGTGTGTGTGTGTGT 26585
QY 539 ----- 539
Db 26586 GGCGGTGTTCGCGAAGACTCCCGCGCGCTGTGACGCCAGATCGGCGAGCTCGCCCGCT 26645
QY 540 -----AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerG 555
Db 26646 AC CGGAGCGGTGTGTACGAGCTGTGCGCGGTGTGCGCGCGCGCTGTGTGTGTGTGTGT 26705
QY 555 lyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerG 575
Db 26706 GTACGCGGATGGAGCACCGCGCGGTGTGCGGTGTGCGGACACCGCGGAGGCACTGCGGAGC 26765
QY 575 lnglyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluA 595
Db 26766 CCTCGGATGCGCGAAGGACTGTGACGCGCGCTGTGCGGCGCGGTGTGCGGTGTGCGGT 26825
QY 595 laIleThrProGlnGlnSerAlaGluLeuAlaSerMet-----GluLeuAsnGluSerS 613
Db 26826 TCCTCTTCCCGCGCAGGCGACGCGGTGTGCGCGGTGTGCGGCGCGGCACTCTCTGTGACAGCT 26885
QY 613 er-----ArgTyrThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluH 630
Db 26886 CACCGGAGTTCTGTGCTGTGCTGTGCGGCGAATGCGAGACCGCG-----CTCTCCC 26933
QY 630 isGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysL 650
Db 26934 GCTACGTGCGACTGTCT----- 26950
QY 650 ysAsnPheTyrPheAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisL 670
Db 26951 -----CTTGAAGCGCTGTGCTCGACAG--- 26971
QY 670 ysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaA 690
Db 26972 -----GAACC-CGCGCGACCCACGCTGCGCGGTGTGCGGTGTGCGGTGTGCGGT 27015
QY 690 laSerGluGluAlaAlaPheProValValGluAspGluGluMetGluAlaSerGlyV 710
Db 27016 -----CGTGACCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 27037
QY 710 alSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnG 730
Db 27038 TCGCTGGC-----GAAGGTCTGGCA-----GCACCGCGCATCACCCCGCGCGCTC 27085
QY 730 luValPro-----ArgGlyGluCysSerGlyProAlaT 741
Db 27086 GTCGGCGCATCGCAGGCGGAGATCGCGCGGTGTGCGGTGTGCGGTGTGCGGTGTGCGGT 27144
QY 741 hrVal---AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAla 760
Db 27145 CGACGCGCGCGGT 27204
QY 760 ys-----AspThrG 763
Db 27205 GGCGCGCATGTCTCTCTCGCCCTCGAGGCGCGGTCTCTGAAGCGAGCTGTGAGCGACTT 27264
QY 763 lyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProProGlyProp 783
Db 27265 CGACGCGACTCTCGGT 27318
QY 783 roThrProProArg---ArgThrSerArgAlaProIleGluProThr----- 797
Db 27319 CGACCGACCCAGATCGAGAACTCGCGCGCATCTGGAGGCGCGAGCGGTGTGTGTGTGTGTGT 27378

producing desosamine transferase which transfers desosamine to substrate polyketides.

Claim 2; Col 21-54; 132pp; English.

The present invention relates to recombinant DNA compounds that encode *Streptomyces venezuelae* (Sv) narbonolide polyketide synthases (PKSs). The recombinant PKSs are derived from narbonolide PKS and other genes involved in narbomycin and picromycin biosynthesis in recombinant host cells. The invention also discloses the *S. venezuelae* PKS gene cluster that results in the production of picromycin. Also disclosed are enzymes such as those responsible for glycosylation and hydroxylation, (e.g. C12 hydroxylase (Pick)), desosamine biosynthesis, and desosaminyl transferase enzymes. The recombinant narbonolide, narbonolide derivatives, and polyketides are useful as antibiotics and as intermediates in the synthesis of compounds for pharmaceutical applications. The present sequence represents cosmid pKOS023-27 containing *S. venezuelae* PKS gene cluster.

Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 U; 0 Other;

```

53  LeuSerProGlySerIlelleGlnProGlnAArgArgProSerLeuLeuSerGluPhe 72
24888  CTCGCCCCCGCGCGCGCGCGCTCCGGAGCT- CGCCGCCCGGCTCAGGGCGCTGCGCGC 24946
73  GlnProGlyAsnGluArgSerGlnGlu-----LeuHisLeuArgPro 86
24947  GGCCCTGGGGGACGACGGCGACGACGCCACCGACCTTGGACGAGCGCTCCGACGACGACCT 25006
87  GluSerHisSerTyrlLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
25007  -----CTTCTCTTTCATCGACAAGAGCTGGCGAGCTTCGACTTCTGACCTGCGCG 25057
107  ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu----- 123
25058  ACACCA-----CGGACACCAACCGGCCACCGACCCCTTACACACGGA 25102
124  -----LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138
25103  ACACGGAACGACAGCGGAGACGGGACCATGGCAACAACGAAGACAAGCTCCGCGAC 25162
139  -----ArgSerLeuThr 142
25163  TACCTCAAGCGCGTCAACGCGGAGCTGCAGAGAACACAGGCGCTGCGCGAGATCGAG 25222
143  GlyLysLeu---GluProValSer-----Pro 150
25223  GGACGACGACGACGACGCGTGGCGATCGTGGCGATGCGCTGCCGCTGCCGGCGGTGTC 25282
151  ProSerProProHisThr-----AspProGluLeuGluLeu---Val 163
25283  GCCTCGCCCGAGGACCTGTGGCAGCTGGTCCCGGGGACGGGGACGC- GATCTCGGAGTT 25341
164  Pro-----ProArgLeuSerLysGluLeuLeuGln---AsnMetAspArgValAsp 180
25342  CCGCGAGGACCGCGCTGGGACGTGGAGGGGCTGTACGACCCCGACCGGACGGGTCCGG 25401
181  ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLys---LysGlnGlnGln 199
25402  CAGGACGCTACTCGCGTCCGCGGATTCCTGCACGACGCGCGGAGTTTCGACGCGGACTT 25461
200  LeuGluGluGluAlaAlaLysProPro----- 208

```

200 LeuGluGluGluAlaAlaLysProPro
||| :: |||||:: |||


```
QY 1798 ----- 1798
Db 31321 GGAGTCCCTGCGTGGAGCGCGCGCGTGGTGGTGGCGGTGTCGC 31380
QY 1799 ----- AspArgAspArgGluArgAspArg-GI 1809
Db 31381 GAAGACTCGCGCGCATGTCAGCGCGCGAGTTCGGCGCATACCGGAAGACCG 31440
QY 1809 uArgGlu-----LysSerIleLeuThrSerThrThrThrValGI 1822
Db 31441 CACGAGCGTGGATCGCGCGTGGCGCGCGCGTGGTGCAGACCGCGTACGGCGATGA 31500
QY 1822 uHisAlaProIle-----TrpArgProGlyThrGluGlnSerSerGlySerSe 1838
Db 31501 GCACCGCGCGTGGTGGCGCGCGAGACCGCGGA-GGCACCTGCGGGACCGCCCTCGGATGC 31559
QY 1838 xGlySerSerGly-----GlyGlyGlyGlySerSerSerArgPr 1851
Db 31560 CGGAAGGACTGGTACGGGGCACGGTACCGATCCGGGCCGGGTGGGT-TCGTCTTCCCC 31618
QY 1851 oAla-----SerHisSerHisAlaHisGlnHisSe 1861
Db 31619 GGCAGGACCGCAGTGGCGCGCATGGCGCGCACTCTCGACAGCTACCCCGAATTC 31678
QY 1861 r-----ProIleSerProArgThrGlnAspAlaLeu----- 1871
Db 31679 GCGCGCGCATGGCCGAATCGAGACCGCACTCTCCCGTACGTCGACTGCTCTCGAA 31738
QY 1872 -----GlnGlnArgProSerValLeuHisAsnThrGlyMetIlyseIleleTh 1888
Db 31739 GCGGTGCTCGCAGCAGGTCCAGCGCACCGACACTCGACCGTGCAGCTGTCGACGCC 31798
QY 1888 xAlaValGluProSer---LysProThrValLeuArgSerThrSerThrSerProVa 1907
Db 31799 GTCACTTCGCGCTCATGGTCTCCCTCGCCCAAGGTCTGGCAGCAGCACCGCATCACCCC 31858
QY 1907 lArgProAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGI 1927
Db 31859 GAGCGCGTCACTCGGCCACTCCAGGCGAGATCG----- 31892
QY 1927 yValTyProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAr 1947
Db 31893 -----CCGCGCGTACGTGCGCGGTGCCCTCACCCTCGACAGCGCGTCTGTGTGTG 31945
QY 1947 gProGlu-----ArgProArgAlaAspThrGlyHisAlaPheLeuAlalysProAl 1965
Db 31946 ACCCTCGCAGCAAGTCCATCGCGCGCCACCTCGCGCGCAAGGCGCGCATGATCCCTC 32005
QY 1965 aArgSerGlyLeuGluProAlaSerSerProSerLys-----GlySerGI 1980
Db 32006 GCCTCAGCGAGGAAGCCACCGCGCAGCGCATCGAGAACCTCCACGGACTGTGCGATGCC 32065
QY 1980 uProArgProLeuValProValSerGlyHisAlaThrIleAlaArgThrProAlaLy 2000
Db 32066 GCGGTCAACGGGCTACCGCCACCGTGTTCGGGCGACCCCAACCCAG-----TCCAA 32119
QY 2000 sAsnLeu-----AlaProHisHisAlaSerProAspProAl 2013
Db 32120 GAACCTGCTCAGGCGTGTAGGCGCGAGCGCATCCGCGCAGGATCATCCCGTCGACTAC 32179
QY 2013 aProProAlaSerAlaSerAsp-ProHisArgGluLysThrGlnSerLysProPheSerI 2033
Db 32180 GCCTCCACAGCGCCCACTCGAGACCATCGAAGCAACTCGCGCACGTCTCTGGCGGG 32239
QY 2033 leGlnGluLeuGluArgSerLeuGlyTyHisGlySerSerTySerProGluGly 2053
Db 32240 TT----- 32242
QY 2053 alGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysH 2073
Db 32243 TCCCCCAGACACCCCA-----GGTCCCTCTTCTCCAC----- 32277
QY 2073 isLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys----- 2089
```

```
Db 32278 -----CCTCGAAGGCACCTGTGATCACCGAAACCCGCTCTCGACGGCGGTAC 32323
QY 2090 -----GlnProGlyPro-----ValLysLeuGly---GlyGluAlaHisL 2103
Db 32324 TGTATCGGACAACTCCGCATCTGTGGGTTCGCGCGCGCGCTCGAGACCTCGCCACC 32383
QY 2103 euProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrA 2123
Db 32384 GACGAAGGCTTACCCACCTTCATCGAGGTGAGCGCCACCCCTCTCT-----CACCATG 32437
QY 2123 laProGlyValLysGlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValI 2143
Db 32438 ACCCTCCCGACAAGTCAACCG-----CCTGGCCACCC-TCGACACCGCA 32481
QY 2143 leThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProL 2163
Db 32482 GACGCGGAGACACACCGCC--TCACACCTCTCTTCGAGGCTTGGCGCAACGCGCT 32538
QY 2163 euTySerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspL 2183
Db 32539 CGCCCTCGACTGGCGCTCCCTCTGCGCGCACCGGCGCCTCAGCCCGCGGTCC--- 32594
QY 2183 euTyLeuPro-ProProAspHisGlyAlaProAlaArgGlySerPro----- 2198
Db 32595 -----CGACCTCCGACGTACGCTTCCAGCA-----CGCTCGTACTGGAT 32637
QY 2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly-GI 2218
Db 32638 CACCCCGCGGTGCGCGGAGCGCGCGCGCACACCGCTTCGCGGCGCGGCGCTCGC 32697
QY 2218 yGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerAr 2238
Db 32698 CGAGACGGGCTCGCTGGGCGCGCGTGCAGGAGCTCGACGAGGAGCGCG---CG 32754
QY 2238 gSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGI 2258
Db 32755 CAGCGCGTACTCGCGATGGTATGCGGCA-GGCGGCGCTCCGTCTCGTGGCGACTCGC 32813
QY 2258 ySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSe 2278
Db 32814 CCGAAGAGTCCC-----CGTCGACCGCGCGCTGCGGAGATCGGTTTCGACTCGC 32864
QY 2278 rAsnSerAlaMetValLysSerLysGlnGluIleAsnLysLys---LeuAsnThrHi 2297
Db 32865 TGACCGCGTGCAGTTCGCGCAACCGGTCAACCGGTGACCGGTCTCCAGTCCGCGCCA 32924
QY 2297 sAsnArg---AsnGluProGluTyArgAsnIleSerGlnProGlyThrGluIlePheAsnMe 2316
Db 32925 CGCTCGTGTTCGAGCACCGCGACCGCGCTCG- CGCTCGCGGAGCGCATCAGCGACGAGCTG 32983
QY 2316 tProAlaIleThrGlyThrGlyLeuMetThrTyArgSer-----GlnAlaValGI 2333
Db 32984 GCCG-----AGCGGAACCTGGGCC-GTCGCGGAGCGCTCGATCAGCAGGCGGAGGA 33036
QY 2333 nGluHisAlaSerThrAsnMetGly----- 2341
Db 33037 GGAGAAGCGCGCTCGCGCGGCGCGCTCGCGGCGCGACACCGCGCGCGCGCGCG 33096
QY 2342 -----LeuGluAlaIleArgLysAlaLeuMetGlyLysTyArgProGluGluSe 2360
Db 33097 GATGTTCCGCGCGCTGTTCGCGGAGCGCGTG-----GAGGACGACCGGTACGCGGAGTT 33150
QY 2360 rProProLeu-----SerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSer----- 2376
Db 33151 CCTCGAGGTCTCGCGAAGCTTCGCGTTCGCGCGCGCTTCGCTTCGCGCGGCGCTG 33210
QY 2377 -----LeuProAlaAlaMetProIleThrAlaAlaAspGlyAr 2389
Db 33211 CTCGAGCGGCTCGACCGGTGTCTGCTCGCGCGGTTCGACGACCGCGCGGAGGCGC 33270
QY 2389 gSer-----AspHisThrLeuThrSerProGlyGly----- 2400
```

29265 ACGCCGCGTGGCGCCGACCGCGCAGTCTGCTGGCGCTGGCGCGATCATCCGCGCTCCCG 29324
 Db
 1231 -----HisGlyThrProAlaAspVal-----LeuTyrIysGlyThrIleThrArgIleIleG 1248
 Qy
 29325 AGCCGAGCCGCGCGCGCTCTCTACCTCTGCTCCGATCCACCGCGCGCGTACTCG 29384
 Db
 1248 lYgluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisV 1268
 Qy
 29385 GCCATTCTCCCGCGCGGTGGCGCCCGCGCGCTGCTTTCACC-----GAGCTCGGTTCCG 29441
 Db
 1268 alIeTyrGluGlyIysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrG 1288
 Qy
 29442 A-CTCGTGAAGCGCGTGGCGCTCCGCAACAGCTCTCCAGTGTGTGCGCAA-----C 29494
 Db
 1288 lncYserLysGluAspGly-----ArgSerSerSerGlyProp 1301
 Qy
 29495 AGGTCTCCCGCCACACCGTCTTCGACACCCGAGCGCGCGCACTCGCGCGCACCTC 29554
 Db
 1301 roHis-----GluThr-AlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
 Qy
 29555 CACGAGCGGTACTCTCGACCGCGCGAGCGCGCGCGAG-----GACTGGAGGGCGG 29608
 Db
 1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
 Qy
 29609 GTGCGCGCGCGCTGGCGCGAACTGCGCTCGACCGGTGCG-GGACGGGGGTCTTCGA 29667
 Db
 1338 GluArgHisSerProHisHisLeuLysGluGlnHisIleArg-GlySerIleThrG 1357
 Qy
 29668 CACCGTCTCGCGCTCACCGCATCGAGCGCGCGGTTCGCGCGGTTCGAGCGCGG 29727
 Db
 1357 nGlyle---ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLy 1376
 Qy
 29728 CGCCCGCGACCTGTGTGGAGCGCGAGCGGTGATGATGACACTGGACGCGGAGCGCT 29787
 Db
 1376 sleuLeuLysArgGluGly----- 1382
 Qy
 29788 GATCGGATGGTCTCGCGCCCGGTAAACACTGACCGCGCGTCTGCGCCCGACGCGC 29847
 Db
 1383 -----ThrProProProProProSe 1390
 Qy
 29848 CGCACCCCGCGCATCCCGCGCACCCCGCGCCCGCACACCGCCATCCACGAGC 29907
 Db
 1390 rArgAspLeuThrGluAlaTyrLysThr-----GlnAlaLeu-GlyProL 1405
 Qy
 29908 GGAAGACCAACCCAGATACAGTTCACAGAACAGTGTGGAGCGTCTGCGCGCTC 29967
 Db
 1405 eulysLeuLysProAlaHisGluGlyLeuValAlaThrVal-----LysG 1420
 Qy
 29968 TCTCAAGGAGAGCAAGAACTCCGGAAGAGAGCGTCTGCGCGCGCGACCGTCCGCGAGGA 30027
 Db
 1420 luAlaGlyArgSerIle-HisGluIleProArgGluGluLeuArgHisThrProGluLeu 1439
 Qy
 30028 GCCATGGCGATGTCGCGCATGAGTGTGCGGTTCGGCGCGCGAATCCCGTCTCCCGAGGA 30087
 Db
 1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr----- 1455
 Qy
 30088 CCTCTGGAGCGCTCGCGCGCGGCAAGCACTGTGTCGAGGTACCGAGAGCGCGG 30147
 Db
 1456 -----ProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLys 1470
 Qy
 30148 CTGGGACATCGACTCTCTACGA-----CCGGTGGCGCGCGCGCAAGGG----- 30192
 Db
 1471 HisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeu 1490
 Qy
 30193 CACGAGTACGT-----CCGCAACGCGCGGTCTCTCG-ACGACGCGCGCG 30236
 Db
 1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGlySerLeuLysSer 1510
 Qy
 30237 GATTTCGAGCGCGCTTCTTCGGGATCTCGCGCGCGCGAGCGCTTCGCCATGAGCCCGCAGC 30296
 Db
 1511 Arg-----ProGlyThrAlaSerSerSerGlyGlySerIleAlaArg 1524
 Qy
 30297 AGCGGCAGCTCTCTGAAGACCTCTTGGGAGGTCTTCGAGCGCGCGCGCATCGACCCCGCT 30356
 Db
 1524 ----- 1524
 Qy
 30357 CGTCCGCGCGCACCGACGTCGCGGTAGTGGGTGGCTTACCAGGACTACGCGCGG 30416
 Db
 1525 -----GlyAlaProValIleValProGluLeuGlyLysPro----- 1536
 Qy
 30417 ACATCCGGGTGCCGCCCGCGGACCGCGGTACGTCTCACCGGCAACTCTCCCGCGG 30476
 Db
 1537 -----ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAla 1551
 Qy
 30477 TGGCCTCCGCGCGCATCGGTACTCTCCGCTCGCTCGCTCGCGACCGCGCTGACCGTGACA 30536
 Db
 1552 GlyHisLeuPro-----ArgGlySerProValThr----- 1561
 Qy
 30537 CGCGGTGCTCTCTCTGCTCGCTCGCTCGCTCGCGACCGCGCGCGGTTC----- 30596
 Db
 1562 -----MetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
 Qy
 30597 ACTGCTCGACGCACTCGTGGCGCGGTCTCTCGCGACCGCGCGCGGTTC----- 30653
 Db
 1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
 Qy
 30654 TCAGTTCAGACGCGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30704
 Db
 1595 SerPro-----HisSerThrValProGluHisHisProHisProLysSer-ProTyrGl 1612
 Qy
 30705 TCGCTCGCGCGCGCGCGCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30764
 Db
 1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
 Qy
 30765 GGCTCTCCGACG 30789
 Db
 1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuPr 1652
 Qy
 30790 -----CCACCGGT-----CCTGGCGCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 30830
 Db
 1652 oA-rgHisLeuAlaProAsnPro-ThrTyrProHisIleuTyrProTyrLeuIleArgG 1672
 Qy
 30831 AGGACGCGCGGAGCAAGCGCTCACGCGTCCGCGAC----- 30865
 Db
 1672 lYtyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrS 1692
 Qy
 30866 --GGGCGCTCCGAGCGCGCTG--ATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30920
 Db
 1692 erGlnGlnMetHis-----HisAsnThrAlaThr----- 1701
 Qy
 30921 CGAGCGAGTGGAGCTGCTGGAGGCGCACGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30980
 Db
 1702 --AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro-----ArgG 1717
 Qy
 30981 AGGCGCGCGCGCTGCTCGCGCACGTACCGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31040
 Db
 1717 luSerSerLeuAlaLeuAsnTyr-----AlaAlaGlyProArgGlyIleI 1732
 Qy
 31041 TGGGCGCGCTGAAGTCCGACATCGGCGCACCGAGGCGCGCTTCGGGTGTGCGCGGTGTCA 31100
 Db
 1732 leAspLeuSerGlnValProHisLeuProValLeuValProPro-ThrProGlyThrPro 1751
 Qy
 31101 TCAAGATGTTCAGCGCGTGGCG 31154
 Db
 1752 AlaThrAlaMetAspArgLeuAlaTyrLeuProThr-----AlaProGlnPro 1767
 Qy
 31155 ACGAGCCGACCGACGAGTTCGCTGGTGGCGCGGTTCGGTTCGCTAGCTGTCTACCGCGCGCG 31214
 Db
 1768 PheSerSerArgHisSerSerProLeuSerProGlyGlyProThrHisLeuThrLys 1787
 Qy
 31215 TGGACTGCCGGA--GCG 31272
 Db
 1788 ProThrThrThrSerSerSerGluArg----- 1798
 Qy
 31273 -----GGCGCGGACGACGCGCAGTCTGCTGGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCG 31320
 Db

```
QY 760 ys-----AspThrG 763
Db 27205 GGGGGCATGATCTCCCTCCCTCGACGAGGGCGCTCTCTGAAGCGACTGAGCGACTT 27264
QY 763 lYGlnHenglyProlyProAlaThrLeuGlyAlaAspGlyProProGlyProP 763
Db 27265 CGAGGACTCTCGTCCCGCGCTCAACG-----GCCACCGCCACCGTCTCTCCGG 27318
QY 783 roThrProProArg---ArgThrSerArgAlaProIleGluProThr----- 797
Db 27319 CGACCCGACCCAGATCGAGAACTCGCCCGCACTCGCGAGCGCGCGCTCGCGCG 27378
QY 797 ----- 797
Db 27379 GATCATCCCGTTCGACTACGCTCCACAGCCGCGAGTTCGAGATCATCGAGAAGAGCT 27438
QY 798 -----ProAlaS 800
Db 27439 GGGCGAGGTCTCGCGGACTCGCCCGGAGGTTCGCGAGCTCGCAGCTGCCGTCTTCTCCACCT 27498
QY 800 erGluAlaThrGlyAlaPro----- 806
Db 27499 CGAAGCACCTGGATCACCGAGCGGTCTCGACGCGACCTACTGTACCGCAACCTGCG 27558
QY 807 -----ThrProProAlaPro----- 813
Db 27559 CCATCGGTGGGTTCGCCCGCGCGTGGAGACCTTGGCGGTTCACGGCTTCCACCCACTT 27618
QY 814 --SerProSerAlaProProValProValProlyGluGluGluGluGluThrAlaA 833
Db 27619 CATCGAGGTGAGCGCCACCGCGCTCTCACATGACCTCCCGAGACCGTCAACCGGCT 27678
QY 833 laAlaProProValGluGluGluGluGln-----LysProProA 847
Db 27679 CGGCACCTCCGCGCGAAACAGGAGCGAGGTCTGGTCACTCTCGCCGAGC 27738
QY 847 laAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysT 867
Db 27739 CTGGGCAACGGGTCTACCATGACTGCGGCGCCATCTCTCCCAACCGCGCCACCA 27798
QY 867 hrGluGluAlaGluGluGlyProAlaLys----- 876
Db 27799 CCGGAGCTCCCACTACCGCTTCAGACCGAGCGCTTCTGGTGCAGAGCTCCGCGCC 27858
QY 877 -----GlyLysAspAlaGluAla-----A 883
Db 27859 CACCAGCGCGCGACGACTGGGTTCACCGGTTCGAGTGAAGCGCTGAGCGCTCCGG 27918
QY 883 laGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySer----- 900
Db 27919 CCAGCGGACCTGTCCGGG-----GGTGGATCGTCCGCGT 27954
QY 900 ----- 900
Db 27955 CGGAGCGAGCCAGACCGAGCTGTGGCGCGCTGAAGCGCGCGGAGCGGAGTTCGA 28014
QY 901 -----GlyValAlaThrThrAlaLysSerSer-----GlyAlaProGln- 913
Db 28015 CGTACTGGAAGCGGGCGGACGACCGTTCGAGCGCTCGCGCGCGGTTCACCGCACT 28074
QY 914 -----AspSerAspSerSerAlaThrCys----- 921
Db 28075 GAGCACCAGCGAGCGCTTCACCGCGGTTCGCTCGTTCGAGCACTCTGCGCACAGGT 28134
QY 922 -----SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnA 936
Db 28135 CGCTGGGTGAGCACTCGCGGACCGCGGAATCAAGCGCGCCCTGTGGTCCGTTCACCCA 28194
QY 936 rgLeuLeuSerProArg-----ProSerLeuLeuThrProThrGlyAspP 951
Db 28195 GGGCGCGGTCTCTCGTCGAGCTCTCGACACCCCGCG-----ACCCGACCGGGCCAT 28248
QY 951 roArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnAlaA 971
28249 GCTCTGGGCGCTCGGCGCGGTCTGTCCTTGAGCAACCCGACCGCTGGGCGCGCTCGT 28308
QY 971 laAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlap 991
Db 28309 CGACTCCCGCGCCAGCCCGATG-----CCGCGCGCTCGGCCACCTCTGTAC 28356
QY 991 roThrLysProAlaProAla-----ProProProProGlnAsnLeu- 1005
Db 28357 CGCACTCTCGGCGCCACCGCGGAGGACGATCGCATCGCACCAACCGACTCCACGC 28416
QY 1006 -----GlnProGluSerAspAlaProG 1013
Db 28417 CGCCCGCTCGCCCGCGCACCCCTCCACGGAAGTCCGCCACCCGAGCTGGCAGCCCA 28476
QY 1013 lnglnProGlySerSerProArgGlyLysSerArgSerProAlaProPro----- 1029
Db 28477 CGGCACCGTCTCATCACCGCGGACCGGAGCCCTCGGCGAGCACCGCGCTGGAT 28536
QY 1030 -----AlaAspLysGluAlaAlaAlaGluAlaGlnLysLeuProGlyA 1045
Db 28537 GGGCCACCGAGCGGAAACACCTCTCTCTGTCTGTCAGCGCGGCGAAACAGCCCGG 28596
QY 1045 spProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysA 1065
Db 28597 AGCCACCC-----AACTCACCGCCGAACTCACCGCATCGGCGCGCCGCTCAC 28644
QY 1065 laserProHisAlaProAspProSerAlaPheSerTyrrAlaProGlyHisProleup 1085
Db 28645 CATCGCGCTGGAGCTCGCGACCCCGCACCGCATCGCACCTCTCGACCGCATCCC 28704
QY 1085 roLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnP 1105
Db 28705 CG-----CCGAGACGCCCTCACGCGC-----TCGTCCA 28734
QY 1105 roProProLeuLysSerSerAla-----LysHisProS 1116
Db 28735 CACCGCGCGCGCACCGCGCGGATCCGTGGAGTCAACCGGCCCGGAGGACATCGCCG 28794
QY 1116 erValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValP 1136
Db 28795 CATCTGGGCGGAGACGAGCGCGCGCGCTCTCGAGCACTCTCGCGGCGACTCC 28854
QY 1136 roTyrrSerGluHisAlaLysAlaProValGlyPro-----V 1148
Db 28855 GCTGGAGCGCTCTGCTCTCTCTCTCGAAACCGCGGCTCTGGGCGAGCGGAGCCAGG 28914
QY 1148 alThrMetGlyLeuProLeuPro-----MetAspProLysLysLeuAlaProPheS 1165
Db 28915 CGTCTACGCGCGCGCAACGCCACCTCGACCGCTCGCGCGCGCGCGCGCGCC----- 28970
QY 1165 erGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSer--L 1184
Db 28971 --GGGCGGAGACCGCGCACTCTGCTCGCT-----GGGCGCTCTGGGCGCGGAGCGCATGG 29025
QY 1184 euGlyValProThrAlaGlnGluAlaSerValLeu----- 1195
Db 29026 CCGGGCGCGCGACGACGCGTACTTGGCAGCGTTCGCGCATCGTCCGATGAGCCCGACCG 29085
QY 1196 -----A 1196
Db 29086 CGCCCTGGACAACTGGCCAAAGCCCTGAGCCACGACGAGACCTCTGCTCGCGTGGCCGA 29145
QY 1196 rgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys----- 1210
Db 29146 TGTGACTGGAGCGGTTCGCG--CCCGCGTTCAGGTGTCCGTCCTCCAGCGCTTCTGCTCG 29204
QY 1211 --GlyIleProSerThrArg-----ValProSerA 1220
Db 29205 ACGGCTCCCGGAGCGCGCGAGCGCTCGCGCACCCGCTCGGTCTGCGCGCTCCCGCG 29264
QY 1220 spSerAlaIleThrTyrrArgGly-----SerIleThr----- 1230
```



```
Qy 2490 ProLeuAlaGly 2493
|||::: |||
Db 33655 CCGTTCGCGG 33666

RESULT 88
AAZ56001
ID AAZ56001 standard; DNA; 38506 BP.
XX
AC AAZ56001;
XX
DT 23-MAR-2000 (first entry)
XX
DE Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.
XX
KW Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;
KW antibiotic production; narbomycin; picromycin; ds.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT 70..13725
CDS /*tag= a
FT /product= "PICAI"
FT /note= "Narbonolide synthase subunit 1"
FT 13830..25049
CDS /*tag= b
FT /product= "PICAI"
FT /note= "Narbonolide synthase subunit 2"
FT 25133..29821
CDS /*tag= c
FT /product= "PICAI"
FT /note= "Narbonolide synthase subunit 3"
FT 29924..33964
CDS /*tag= d
FT /product= "PICAI"
FT /note= "Narbonolide synthase subunit 4"
FT 33961..34806
CDS /*tag= e
FT /product= "PicB"
FT /note= "Contains typeII thioesterase domain"
FT 34863..36011
CDS /*tag= f
FT /product= "PICII"
FT /note= "4-keto-6-deoxyglucose isomerase"
FT 36159..37439
CDS /*tag= g
FT /product= "PICIII"
FT /note= "Desosaminyl transferase"
FT 37529..38242
CDS /*tag= h
FT /product= "PICCVI"
FT /note= "3-amino dimethyltransferase"
XX
WO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US011814.
XX
PR 28-MAY-1998; 98US-0087080P.
PR 28-AUG-1998; 98US-00141908.
PR 22-SEP-1998; 98US-0100880P.
PR 08-FEB-1999; 99US-0119139P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX WPI; 2000-072618/06.
DR P-PSDB: AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
DR AAY67208, AAY67211.
XX
```

New recombinant DNA encoding a domain of narbonolide polyketide synthase, for production of ketolide antibiotics.

Example 2; Page 16-27; 98pp; English.

This is the recombinant cosmid pKOS023-27 DNA sequence which contains a Streptomyces venezuelae DNA insert. The cosmid contains open reading frames which encode the various modules of the narbonolide polyketide synthase (PKS). The invention relates to recombinant DNA containing a coding sequence for a narbonolide PKS. Polyketides are compounds synthesised from 2-carbon units through a series of condensations and subsequent modifications. Modular PKSs are responsible for the production of many antibiotics including picromycin. The narbonolide PKS consists of a loading module, six extender modules, and two thioester domains. Four proteins make up the narbonolide PKS (PICAI, PICAI, PICAI and PICAI). PICAI includes the loading module and extender modules 1 and 2, PICAI includes extender modules 3 and 4, PICAI includes extender modules 5 and 6, PICAI includes extender module 6 and a type II thioesterase domain. The second type II thioesterase domain is found on the PICB protein. The nucleotide sequences encoding all of these proteins can be isolated in recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl transferase enzyme is required for this conversion, and the desosamine DNA of the invention is used to express, in transformed cells, biosynthetic genes are also found in cosmid pKOS023-27. The recombinant narbonolide (or its derivatives) or other ketolides (particularly hybrids), which may then be converted (e.g. by other enzymes recombinantly expressed in the same hosts) to polyketide antibiotics or their intermediates. The antibiotics are useful in human or veterinary medicine

Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 U; 0 Other;

Alignment Scores:

| Pred. No.: | 1-73e-09 | Length: | 38506 |
|------------------------|----------|---------------|-------|
| Score: | 600.00 | Matches: | 697 |
| Percent Similarity: | 31.71% | Conservative: | 322 |
| Best Local Similarity: | 21.69% | Mismatches: | 1150 |
| Query Match: | 4.54% | Indels: | 1061 |
| DB: | 3 | Gaps: | 146 |

US-09-522-753-5 (1-2517) x AAZ56001 (1-38506)

Qy 53 LeuSerProGlySerIleIleGlnProGlnArgArgProSerLeuSerGluPhe 72
|||::: |||
Db 24888 CTGCGCGCGCGCGCGCGCGCTCGGAGCT-CGCGCGCGCTCAGGGCGTGGCGCG 24946

Qy 73 GlnProGlyAsnGluArgSerGlnGlu-----LeuHisLeuArgPro 86
|||::: |||
Db 24947 GGCCCTGGGGGAGCAGCGCGACGCGCCACCTGGAGCGGGCTCGGACGACCT 25006

Qy 87 GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
|||::: |||
Db 25007 -----CTTCTCCTTCATCGACAAGGAGCTGGGGCGACTCCGACTTCTGACCTGCCCG 25057

Qy 107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu----- 123
|||::: |||
Db 25058 ACACCA-----CCGGCACCACCGCGCACCGCCCTTCACACACGGA 25102

Qy 124 -----LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138
|||::: |||
Db 25103 ACACGGAACGACAGCGGAGAACGGGAGCCATGCGCAACACGACGAGCTCCGCGAC 25162

Qy 139 -----ArgSerLeuThr 142
|||::: |||

Db 25163 TACCTCAAGCGCGTACCGCGCGAGTGCAGCAGACACGAGCGCTCTGCGGAGATCGAG 25222

Qy 143 GlyLysLeu---GluProValSer-----Pro 150
|||::: |||

Db 25223 GGACGACGACGACGAGCGCGTGGGATCGTGGGATCGTGGCGCTGCGCGCGGTGTC 25282

Qy 151 ProSerProProHisThr-----AspProGluLeuGluLeu---Val 163
|||::: |||

| | | | | | | | | | |
|----|-------|--|--|-------|-------|--|--|---|-------|
| Db | 31679 | GC | CGCGCCATGCGCGAATCGAGACCGCACTCTCCCGGTACGTCGACTGGTCTCTCGAA | 31738 | Db | 32595 | ----- | CCGACCTCCCGACGTACGCTTCACGA | 32637 |
| Qy | 1872 | ----- | GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh | 1888 | Qy | 2199 | His | SerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly-Gl | 2218 |
| Db | 31739 | GC | CGTCTCGACACAGGTCTCCAGCGCACCGCACTCGACCGCTCGACGTCGTCAGCCCC | 31798 | Db | 32638 | CAG | CCCGCGGGTCCCGCGAGCGCCCGCCACACCGCTTCCGGGCGCGAGCGCTCGC | 32697 |
| Qy | 1888 | rAlaValGluProSer--- | LysProThrValLeuArgSerThrSerThrSerSerProVa | 1907 | Qy | 2218 | yGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerAr | 2238 | |
| Db | 31799 | GT | CACCTTCGCGTCTGCTCTCCGCAAGGTCTGGCAGCACACACGCGCATCACCCCC | 31858 | Db | 32698 | CG | AGCGGGTCTCGTGGCGCGGCTCGGAGACCTCGACGAGGAGCGCG--CG | 32754 |
| Qy | 1907 | lArgProAlaAlaThrPheProAlaThrHisCysProLeuGlyGlyThrLeuAspGl | 1927 | Qy | 2238 | gSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGl | 2258 | | |
| Db | 31859 | GAG | CGCGTCTATCGGCCCATCTCCAGCGCGAGATCG----- | 31892 | Db | 32755 | CAG | CGCGCTACTCGCGATGTGATCGGCA-GGGCGCTCTCGTCTCGGTCGCGATCGC | 32813 |
| Qy | 1927 | yValTyProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAr | 1947 | Qy | 2258 | ySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSe | 2278 | | |
| Db | 31893 | ----- | CCGCGCGGTAGTCCGCGTCCCTCACCTCGACGCGCGCTCGTGTCTG | 31945 | Db | 32814 | CG | AGAGAGTCCC-----CGTCAGCCGCGCTCGGGAGATCGGCTTCGACTCGC | 32864 |
| Qy | 1947 | gProGlu----- | ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAl | 1965 | Qy | 2278 | rAsnSerAlaMetValLysSerLysGlnGlnIleAsnLysLys--- | LeuAsnThrHi | 2297 |
| Db | 31946 | ACC | TCCGACGAAGTCCATCGCGCCACCTCGCGCGCAAGCGCGCATCTCCCTC | 32005 | Db | 32865 | TG | ACCGCGTCTCGACTTCGCAACCGGTCAACCGGTGACCGGTCTCCAGCTGCGCGCCA | 32924 |
| Qy | 1965 | aArgSerGlyLeuGluProAlaSerSerProSerLys----- | GlySerGl | 1980 | Qy | 2297 | sAsnArg--- | AsnGluProGluTyArgAsnLysSerGlnProGlyThrGluIlePheAsnMe | 2316 |
| Db | 32006 | GC | CTCAGCGAGGAGCCACCGCGCAGCGCATCGAGAACCTCCAGGACTGTGATGCC | 32065 | Db | 32925 | CG | TCTGTTTCAGACACCGACCGCGCTCGCGCGAGCGCATCAGCGAGCTG | 32983 |
| Qy | 1980 | uProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLy | 2000 | Qy | 2316 | tProAlaIleThrGlyThrGlyLeuMetThrTyArgSer----- | GlnAlaValGl | 2333 | |
| Db | 32066 | CG | GTCAACGGGCTACCGCCACCGTGTTCGGCGCAGCCACCCAGA----- | 32119 | Db | 32984 | GCG | -----ACGGGACTGGGCC-GTCGCCGAGCGCTGGATCAGGACGCGGAGGA | 33036 |
| Qy | 2000 | sAsnLeu----- | AlaProHisHisAlaSerProAspProProAl | 2013 | Qy | 2333 | nGluHisAlaSerThrAsnMetGly----- | ----- | 2341 |
| Db | 32120 | GA | CTTGCTCAGCGGTGAGCGCGACGCGCATCGCGCAGCGATCATCCCGTCGACTAC | 32179 | Db | 33037 | GG | AGAGCGCGCGCTCCCGCGCGCGCTCCCGGGCCGACACCGCGCGCGCGCGG | 33096 |
| Qy | 2013 | aProAlaSerAlaSerAsp-ProHisArgGluLysThrGlnSerLysProPheSerI | 2033 | Qy | 2342 | ----- | LeuGluAlaIleAlaArgLysAlaLeuMetGlyLysTyArgAsnGlnTrpGluGluSe | 2360 | |
| Db | 32180 | GC | CTCCACAGCGCCCGCATCGAGACCATCGAGACGAACTCGCGAGTCTTCGCGGG | 32239 | Db | 33097 | GAT | GTTCGCGCGCTTTCGCGGAGCGCGT-----GAGGACGCGGTTCGCGGAGTT | 33150 |
| Qy | 2033 | leGlnGluLeuArgSerLeuGlyTyHisGlySerSerTySerProGluGlyV | 2053 | Qy | 2360 | rProProLeu----- | SerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSer----- | 2376 | |
| Db | 32240 | TT----- | ----- | 32242 | Db | 33151 | CCT | CGAGTCTCGCGGAAAGCTTCGCGTTCGCGCGAGTTCGCTCGCGAGGCGT | 33210 |
| Qy | 2053 | alGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysH | 2073 | Qy | 2377 | ----- | LeuProAlaAlaMetProIleThrAlaAlaAspGlyAr | 2389 | |
| Db | 32243 | TCC | CCCCAGACACCCCA-----GGTCCCTCTTCTCTCAC----- | 32277 | Db | 33211 | CT | CGAGCGGCTCGACCCCGGTCTGTCGCGCGGTTCGACGACCGCGCGGAGCGCG | 33270 |
| Qy | 2073 | isLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys----- | 2089 | Qy | 2389 | gSer----- | AspHisThrLeuThrSerProGlyGlyGly----- | 2400 | |
| Db | 32278 | ----- | CCTCGAGGCGACCTGGATCACCGAACCCTCGACGCGCGCTAC | 32323 | Db | 33271 | TG | CGTCTCGTGGTGCACCGCGCACCGCGGAAACGCGCGCGAGTTCCTCGC | 33330 |
| Qy | 2090 | ----- | GlnProGlyPro-----ValLysLeuGly-----GlyGluAlaAlaHisL | 2103 | Qy | 2400 | ----- | ----- | 2400 |
| Db | 32324 | TG | TACCGCAACCTCGCCATCGTGTGGCTTCGCGCGCGCGTTCGAGACCTTCGCCACC | 32383 | Db | 33331 | GCT | CAGCACCTCTTCAGGAGGAGCGGAGTTCCTCGCGTACCTCTCCCGGCTACGG | 33390 |
| Qy | 2103 | euProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThra | 2123 | Qy | 2401 | ----- | Gly-LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerP | 2417 | |
| Db | 32384 | GAG | CAAGGCTTCAACCACTTCATCGAGGTACGCGGCCCGCTCT-----CACCATG | 32437 | Db | 33391 | CAC | GGGTACGGGACCGCGCACCGCCCTCTCCCGCGCGATCTCGACACCGCGCTCGACGC | 33450 |
| Qy | 2123 | laProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValI | 2143 | Qy | 2417 | roAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluG | 2437 | | |
| Db | 32438 | ACC | CTCCCGCAAGGTACCGG-----CCTGGCCACCC--TCCGACGCGA | 32481 | Db | 33451 | CC | AGCCCGGCGATCTCTCGGG-----CCGCGCGGCGACCGCCCGTCTCTGTCTCG | 33504 |
| Qy | 2143 | leThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProL | 2163 | Qy | 2437 | ly----- | AspCysAsnArgArg-----ThrProLeuThrAsnArgValT | 2450 | |
| Db | 32482 | GG | ACGCGGACACACCGCC--TCACCACTCTCTCGGAGGCTTGGGCCAACCGGCT | 32538 | Db | 33505 | GC | ATCCCGGCGCGCCCTCTCGCGCACGAGTGGCTTCGCTCGAGCGGGCGACGG | 33564 |
| Qy | 2163 | euTySerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspL | 2183 | Qy | 2450 | rpGluAspArgProSerSerAlaGlySerThrProPheProTyArgAsnProLeuIleMe | 2470 | | |
| Db | 32539 | CG | CCCTTCGACTGGGCTCTCTCTGCGCGCACCGGCGGCTCAGCGCGCGCGTCC-- | 32594 | Db | 33565 | CG | CGCGCGCGCGGATCGTCTGTGTGACCCCT-----ATCCGCC | 33606 |
| Qy | 2183 | euTyLeuPro-ProProAspHisGlyAlaProAlaArgGlySerPro----- | 2198 | Qy | 2470 | rgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeu-ProAlaGlySerGly | 2489 | | |
| | | | | Qy | 33607 | GG----- | GCCATCAGGAGCCCATCGAGGTGTGGAGGAGGAGGAGTGGCGGAGGG | 33654 | |

QY 1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisLeuArg-GlySerIleThrG1 1357
DB 29668 CACCGTCTCGGCTACCGGCATCGAGCCGAGCCGCGGTTCGCGCGGTTCGAGCGCGG 29727
QY 1357 nGlyIle---ProArgSerTyrrValGluAlaGlnGluAspTyrrLeuArgGluAlaLys 1376
DB 29728 CGCGCGCGACCTGTGTGCGAGCGCGGAGCGGTGCATCGACGACCTGGACGCCGAGCGCT 29787
QY 1376 sLeuLeuLysArgGluGly-----1382
DB 29788 GATCCGATGGCTCTCGGCCCCCGTAACACTGACCCGCGCGGTCTCTCCCGACGCGC 29847
QY 1383 -----ThrProProProProProse 1390
DB 29848 CGCACCCCGGCATCCCGCGCACCCACCGCGCCCCACACGCCACACCCCATCCAGCAGC 29907
QY 1390 rArgAspLeuThrGluAlaTyrrLysThr-----GlnAlaLeu-GlyProL 1405
DB 29908 GGAAGACCAACACCCAGATGACGAGTTCACACGACAGTTCGTGTGACGCTCTGCGCGCCTC 29967
QY 1405 eLysLeuLysProAlaHisGluGluLeuValAlaThrVal-----LysG 1420
DB 29968 TCTCAGGAGACGAGAACTCCGAAAGAGACCGCTCGCGCGCGCACCGTCCGCGAGGA 30027
QY 1420 lualaGlyArgSerIle-HisGluIleProArgGluGluLeuArgHisThrProGluLeu 1439
DB 30028 GCCATGGCGATCGTCGCGATGAGTCGCGGTTCGCGGCGGAATCCGCTCCCGCAGGA 30087
QY 1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr-----1455
DB 30088 CCTCTGGGACCGCTCGCCCGCGGCAAGACCTGGTCTCCGAGGTACCCGAGGAGCGCG 30147
QY 1456 -----ProLeuLysTyrrAspThrGlyAlaSerThrThrGlySerLysLys 1470
DB 30148 CTGGGACATCGACTCCCTCTACGA-----CCGCGTCCCGCGGCGCAGG- 30192
QY 1471 HisAspValArgSerLeuIleGlySerProGlyArgThrPheProValHisProLeu 1490
DB 30193 CACGACGTACGT-----CCGCAACCGCGCGTTCCTCG-ACGACGCGCGC 30236
QY 1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrrGluGluSerLeuLysSer 1510
DB 30237 GATTCGACGCGGCTTCTTCGGGATCTCGCGCGGAGGCGCCCTCGCCATGACCCGCGC 30296
QY 1511 Arg-----ProGlyThrAlaSerSerSerGlyGlySerIleAlaArg 1524
DB 30297 AGCGGCAGCTCTCGAAGCTCTCTGGAGGTCTTCGAGCGGCGCGCATCGACCCCGCT 30356
QY 1524 -----1524
DB 30357 CGGTCCGCGCACCGAGTCCGCGGTGTACGTGGGCTGTGCTACGAGACTACGCGCGG 30416
QY 1525 -----GlyAlaProValIleValProGluLeuGlyLysPro-----1536
DB 30417 ACATCCGGGTGCGCCCCGAGAGCACCGCGGTTACGTCGTCACCGCGCACTCTCCGCGC 30476
QY 1537 -----ArgGlnSerProLeuThrTyrrGluAlaAspHisGlyAlaProPheAla 1551
DB 30477 TGGCTCCGGGCGCATCGGTACTCTCCCTCGGCTTGAGGACCCGCGGTACCGTGGACA 30536
QY 1552 GlyHisLeuPro---ArgGlySerProValThr-----1561
DB 30537 CGGCGTCTCTCTTCGCTCGTCCGCTGCACCTCGCCCTGAAAGGGCTCCGGAACGCGC 30596
QY 1562 -----MetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
DB 30597 ACTGCTCGACCGCACTCGTGGCGCGGTGCGCTCTCCGCGACCGCGCGCGCTTCA--- 30653
QY 1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
DB 30654 TCGAGTTCAGCAGCAGCGCATGCGCCCGCAGCGCCGCGCAGC-----GCT 30704

QY 1595 SerPro-----HisSerThrValProGluHisHisProHisProIleSer-ProTyrrG1 1612
DB 30705 TCGCCTCGCGCGGAGCGGCTCGCTCGCGCGAGGCGGTTCGCTCTCTCTCGAAC 30764
QY 1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrrArgSerHisIleProLeuAlaPh 1632
DB 30765 GGCTCTCCAGCAGCGCGCGCGCAAGG-----30789
QY 1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrrLeuPr 1652
DB 30790 -----CCACCGGT---CCTGCGCGTCTGCGCGGCGCGCATCAACC 30830
QY 1652 oArgHisLeuAlaProAsnPro-ThrTyrrProHisLeuTyrrProTyrrLeuIleArgG 1672
DB 30831 AGGACGCGCGAGCAACCGCTCACGGCTCCGCAC-----30865
QY 1672 lyTyrrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrrIleThrS 1692
DB 30866 --GGGCGCTCCCGAGCAGCGCTG---ATCGCGCAGCGCTTGGCGGACGCGCGCTCACGT 30920
QY 1692 erGlnGlnMethis-----HisAsnThrAlaThr-----1701
DB 30921 CGAGCCAGCTGGACGCTCTGGAGGCGCACGGCAGGGGCCCTCTCTCGCGACCCGATCG 30980
QY 1702 --AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro-----ArgG 1717
DB 30981 AGCGCAGCGCTGTCTCGCACGATACGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 31040
QY 1717 luserSerLeuAlaLeuAsnTyrr-----AlaAlaGlyProArgGlyIleI 1732
DB 31041 TGGGAGCGCTGAAGTCAAGTCAATCGGCGCACCGCAGCGCGCTTCGGGTGTCCGCGGTCA 31100
QY 1732 leAspLeuSerGlnValProHisLeuProValLeuValProPro-ThrProGlyThrPro 1751
DB 31101 TCAAGATGTGCGCGCGCTCGCGCACCGGCGGTGTG-----CCGAGACCTCTCACGCGT 31154
QY 1752 AlaThrAlaMetAspArgLeuAlaTyrrLeuProThr-----AlaProGlnPro 1767
DB 31155 ACGACGCGAGCAGCACCGAGTCTGCTCGCGCGCTTCGGTTCGAGTCTCACCGAGCGC 31214
QY 1768 PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLys 1787
DB 31215 TGACGTGGCGCGA--CGCGCGCGCGCGCTCCCGCGCGCGCGCGCTCTCCGCGTTCGGCGT 31272
QY 1788 ProThrThrThrSerSerSerGluArgGluArg-----1798
DB 31273 -----GGCGCGGACGAAACCGCGCACGTCGTCCTCGAGGAGGCGCGCGCGTCA 31320
QY 1798 -----1798
DB 31321 GGAGTCCCTCTCGCTCGAGCGCGCGCGTGGCGCGTGGTGGCGGTGTCTCCG 31380
QY 1799 -----AspArgAspArgGluArgAspArgAspArg-G1 1809
DB 31381 GAAGACTCGCGCGCATCTGGACGCCAGATCGGCGAGCTCGCGCATACCGGAAGACCG 31440
QY 1809 uArgGlu-----LysSerIleLeuThrSerThrThrThrValG1 1822
DB 31441 CACGAGCTGGATCCGCGGTGCGCGCGCGCGCTGGTTCGACAGCGCTACCGCGATGA 31500
QY 1822 uHisAlaProIle-----TrpArgProGlyThrGluGlnSerSerGlySerSe 1838
DB 31501 GCACCGCGCGGTCTCGGTCTGGCGACAGCGCGGA--GGCACTTCGCGGACGCGCTCGGATG 31559
QY 1838 rGlySerSerGly-----GlyGlyGlyGlySerSerSerArgPr 1851
DB 31560 CGAAGACTGGTACCGGGGACCGGTCAACGATCCGCGCGCGGTGGGT-TCGTCTTCCCC 31618
QY 1851 oAla-----SerHisSerHisAlaHisGlnHisSe 1861
DB 31619 GGCGGCGCACGAGTGGCGCGCATGGGCGCGCGCACTCTCTCGACAGCTCACCGAATTC 31678
QY 1861 r-----ProIleSerProArgThrGlnAspAlaLeu-----1871


```
QY 250 ProGlnValGluLeuProLeuTyArgAsnGlnProSerAspThrArgGlnTyrHisGluAsn 269
DB 25702 CCGTATCGCGTACGTCCTCGGTACGAGCGGACCGGCGCTGACCGTGGACACGCGCTGCTC 25761
QY 270 IleLysIleAsnGlnAlaMetArgLysLysLeuLeuLeuTyrPheLysArgArgAsnHis 289
DB 25762 GTCTCGCTGTCGCGCTCGACCTCGCGTGCAGGCGCTCGCAAGGCGAGTGCACAT 25821
QY 290 AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309
DB 25822 GGCCTTCGCGCGTGGTGTACGCTCATGCCAAGC----- 25857
QY 310 GluLysLysValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysVal 329
DB 25858 ---GGACCTGCTGTCGAGTTGAG-----CGGCGAGCGGCGCTGCGCGCGAGCG--- 25905
QY 330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg 349
DB 25906 -----CCGTCGAAGGCGTTCCGCCACCTCGCGGAGCG 25938
QY 350 MetGlnSerArgValGlyGlnArg-----GlySerGlyLeuSerMet-SerAlaAlaAr 367
DB 25939 CTTGCGCGCGCGAGCGCGCGGAGTCTGCTGTGGAGCGCTGTCCGACCGCGCGCG 25998
QY 367 gSerGluHisGluValSerGluIlele-----AspGlyLeuSe 380
DB 25999 CAACGAGACACCGGATCCTCGCGTGTGTCGCGGCGAGCGGTCACACGAGGAGCGCGCAG 26058
QY 380 rGluGlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProMetLeuTy 400
DB 26059 CAACGGCCTCACGGCTCCGACGCGGCGCTCCAGCAGCGCGTTCATCCGACGCGCGCTGCG 26118
QY 400 rAspAla-----AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLe 415
DB 26119 GGACGCGCGGCTCGCGCGGTCAGTGTGAGCGTGTGAGCGCGCAGCGGCGGCGCGG 26178
QY 415 uMetAlaAspProMetLys-----ValTyrLysAsp-ArgGlnValM 429
DB 26179 GCTCGCGGACCGGATCGAGCGCGGCGGCTCATCGCCACCTACGCGGAGGAGAGCAG 26238
QY 429 etAsnMet-----TrpSerGluGlnGluLysGluThrPheArgGluLysPheMetG 446
DB 26239 CGAACAGCGCGCTGAGGTGGCGCGCTTGAAGT-----CGAACATCGCGGCGCAC 26286
QY 446 lnhisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluC 466
DB 26287 GCAGCGCGCGCGGTCGACAGGTGTCAAGATGGTCCAGGCGGATGCGCCACGAGCT 26346
QY 466 ysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgAr 486
DB 26347 GCTGC----- 26351
QY 486 erTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 506
DB 26352 -----CGAAGAGCGTCACGTCGACGAGCGCTCGGACACGATCGACTGTGCGGCGGCGAC 26406
QY 506 lnglnGlnGlnProMetProArgSerSerGlnGlnGluLysAspGluLysGluLysG 526
DB 26407 GGTGGAACTCCTCACCAGCGCGTGTGAGTGGCGG-GAGAAGAGGAGCGCGCGGCTGCGCC 26465
QY 526 lu-----LysGluA 529
DB 26466 GCGCGGCTGCTCTCTCTCGGCATCAGCGGGAGCGAGCGGCGACGTCGTCCTCGAGGAGG 26525
QY 529 lAgLuLysGluGluGluLysProGluValGlu----- 539
DB 26526 CCGCGCGGTCGAGGACTCCCGCGCGCTCGAGCGCGCGCGGTCGCGGTGCGGTGCGCGT 26585
QY 539 ----- 539
DB 26586 GCGCGGTGTCGCGAAGACTCCCGCGCGCTGAGCGCGCGAGATCGGGGAGCTCGCGCGCT 26645
```

```
540 -----AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThr-SerG 555
26646 ACGCGAGCGTCTGTACGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26705
555 lyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerG 575
26706 GTACGCGGATGAGACACCGCGCGGTGCGGTGCGGCGAGCGGAGGAGGACCTCGCGGAGC 26765
575 lnglyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluA 595
26766 CCTCGCGGATCGCGGAGGAGCTGTGTACGCGGCGACGTCCTCGGACGTGGCGCGGTGGCGT 26825
595 laileThrProGlnGlnSerAlaGluLeuAlaSerMet-----GluLeuAsnGluSerS 613
26826 TGTCTCTCCCGCGGAGGCGAGCGAGTGGCGCATGGCGCGCGGCGCGGAACTCTTTGACAGCT 26885
613 er-----ArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluH 630
26886 CACCGGAGTTCGCTGCTCGATGGCGCAATGCGAGACCGCG-----CTCTCCC 26933
630 isGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysL 650
26934 GCTACGTCGACTGCTCT----- 26950
650 ysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisL 670
26951 -----CTTGAAAGCGCTGCTCCGACAG--- 26971
670 ysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaA 690
26972 -----GAACC-CGCGCGACCCACGCTCGCGGTGCGACGTGCTGTCAGGC--- 27015
690 laSerGluGluAlaAlaPheProValValGluAspGluGluMetGluAlaSerGlyV 710
27016 -----CGTGACCTCGCTGTCATGGTC 27037
710 alserGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnG 730
27038 TCGCTGCG-----GAAGGTCTGGCA-----GCACCAGCGCATCACCCCGGCGCGTC 27085
730 luValPro-----ArgGlyGluCysSerGlyProAlaAr 741
27086 GTCGGCCACTCGCAGGCGGAGATGCGCGCGGTAGCTGCGCGGTGCATCTCA-CCCTCGA 27144
741 hrVal---AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAla 760
27145 CGACGCGCGCGCGTCTGTCACCTCGCGCAGCAAGTCTCGCGCGCCACCTCGCGCGCAA 27204
760 ys-----AspThrG 763
27205 GCGCGCATGATCTCCCTCGCCCTCGACGAGCGCGCGCTCTGAAGCGACTGAGCGACTT 27264
763 lyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProGlyProp 783
27265 CGACGAGCTCTCGTTCGCGCGCGTCAACG-----GCCCGACCGCGCACCGTCTCTCCGG 27318
783 roThrProProArg---ArgThrSerArgAlaProIleGluProThr----- 797
27319 CGACCGCGCCAGATCGAGGAAGTCTGCCCGCACCTCGGAGCGCGCGCGCTCGCGCGCG 27378
797 ----- 797
27379 GATCATCCCGGTCTGACTAGCGCTCCACAGCGCGAGGTGAGATCATCGAGAGGAGCT 27438
798 -----ProAlaAs 800
27439 GGCAGAGTCTCTCGCGGAGACTCGCCCGCGCAGGCTCGCGAGTCTTCTTCCACCT 27498
800 erGluAlaThrGlyAlaPro----- 806
27499 CGNAGGCACTGTGATCACCAGCGCGGTGCTCGAGCGGCACTTACTGTGTACCGCACTGCG 27558
807 -----ThrProProAlaProPro----- 813
```



```
Qy 1156 tAspProLysLys-----LeuAlaProPheSerGlyVallysGl 1169
Db 32666 CGACGGCTCGGGAGCGGGGTCTCGACACCGTCTCGCGCTACCGGCATCGAGCC 32725
Qy 1169 nGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAl 1189
Db 32726 CGAGCGG-----GGTTCGGCGGTTCGGACGGCGCGCGCGCCGACCTCGTGC 32773
Qy 1189 aGln---GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySer-- 1207
Db 32774 GGAGCCGAGCGGTTCGATCGA-CGACCTGGACGCGGAGCCCTGATCCGATGGCTCTCG 32832
Qy 1208 -----IleThr-----LysGlyIleProSerThrArgValProSerAspSe 1221
Db 32833 GCCCCGTAACACCTGACCGCGGTCTCGCCCGACCGCGCGACCCCGCGCATCC 32892
Qy 1221 xAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGl 1241
Db 32893 CGGCACACCGCCCGCCACAGCCACA-----ACCCCATCC----- 32929
Qy 1241 yThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAs 1261
Db 32930 -ACGAGCGGAAGACACACCCAGATGACGAGTTCACACGAACAGTTCGTGGAGCTCTGC 32988
Qy 1261 pSerLeuProLysGlyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGl 1281
Db 32989 CGCGCTCTCTCAAGGAGA-ACGAAGAATCCGGAAGAGAGACC---GTCCCGCGCGCCGAC 33044
Qy 1281 uGlyGlyMetSerValThrGlnCysSer-----LysGluAspGlyArgSerSerGl 1299
Db 33045 CGTCGGAGGAGCCATCGGCGATCGTGGCATGAGCTGCGGTTCGCGGGCGGAATCCGG 33104
Qy 1299 yProProHiegluThrAlaAlaProLys-----ArgThrTyrAspMetMetGluGl 1316
Db 33105 TCCCCGAGGACCTCTGGAGCGCGTCCGCGCGGCAAGGACCTGGTCTCGAGGTACCG 33164
Qy 1316 yArgValGlyArgAlaIleSerSerAlaSerIle-GluGly-----LeuMetGlyArgA 1334
Db 33165 GAGGAGCGCGCTCGGACATCGATCCCTCTACGACCGCGTGGCGCGCGGCGGACGCG 33224
Qy 1334 lalleProProGluArgHisSerProHisHisLeuLysGluGlnHisIleArgLys 1354
Db 33225 ACGTACGTCCGACAGCGCGCTTCT-----CGACGACGCGCGCGGATTGACGCG 33275
Qy 1354 erIleThrGlnGlyIle-ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArg 1373
Db 33276 GCCTTCTTCGGGATCTCGCCG-----CGC 33299
Qy 1374 GluAlaLysLeuLeuLysArgGluGlyThrProProProProProSerArgAspLeu 1393
Db 33300 GAGGCCCTCGGCATG-----GACCCGACGACGCGGACGTC 33335
Qy 1394 ThrGluAla-----TyrIleThrGlnAlaLeuGlyProLeuLysLeuLysPro 1409
Db 33336 CTCGAAGCTCTCGGAGGTCTTCGAGCGCGCGCATCGACCGCGCTCGGTCCGCGG 33395
Qy 1410 AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIlePro 1429
Db 33396 ACCGACGCGCGGTATC-----GTGGGCTGTGGCTACGAGAC----- 33434
Qy 1430 ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly 1449
Db 33435 -----TAGCGCGGACATCCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 33476
Qy 1450 SerIleThrGlnGly-ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerIly 1469
Db 33477 TACGTCGTCAACGCACTCTCCCGCGTGGCTCCGCGCGCATCGCTACTCTCCCTCGGC 33536
Qy 1469 slyHisAsp-----ValArgSerLeuIleGlySerProGlyArgThrPh 1484
Db 33537 CTGAGGAGGACCGCGCTGACCGTGGACACGCGGT-----GCTCTCTTCGTCTGTC 33587
```

```
Qy 1484 eProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTy 1504
Db 33588 GCCTGCACCTCGCCC-----TGAGGGCCCTGC-- 33615
Qy 1504 rGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAla-A 1524
Db 33616 -----GMAACGGCGACTGCTCGACCGCACCTCGTGGCGCGC 33650
Qy 1524 rGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyG 1544
Db 33651 GTGGCGTCTCCCGACGCGCGCGGTTCATCGAGTTTCAGCAGCA-----G 33698
Qy 1544 luAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgG 1564
Db 33699 CAGGCATGGCGG-CGACGCGCGGA-----CCAAGGCTTCGCT--CGCGCGCGGA 33748
Qy 1564 luProThrProArgLeuGlnGluGlySerSerSerSerSerLysAlaSerGlnAspArg- 1583
Db 33749 CGGCTCGCTCGGCGGAGGCGGTCTCGCTCTCTCTCGAACGGCTCTCGACGCGG 33808
Qy 1584 -----LysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro- 1601
Db 33809 GCCAAGGCGCACCGGTCTCGGCGTCTCGGCGCGACGCGCATCAACGACGCGCGC 33868
Qy 1602 GluHisHisProHis-----ProIleSerPro---TyrGluHisLeuLeuArgGlyVal-- 1618
Db 33869 GAGCAACGGCTCACGGCTCCGACGCGGCTCCAGCAGCAGCCTGATCCGCGCGCCCT 33928
Qy 1619 -----SerGlyValAspLeuTyrArgSerHisIleProLeuAl 1631
Db 33929 GGCGGACGCGCGCTCACGTGAGCGACGTGGAGCTGTGGAGGCGCAC-- 33977
Qy 1631 aPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLe 1651
Db 33978 -----GGCAGCGGACCGCTCTCGCGACCGCATCGAGCGCGCGGTGCTC-- 34025
Qy 1651 uProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleAr 1671
Db 34026 -----GCCACGTACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 34052
Qy 1671 gGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleTh 1691
Db 34053 -GGCGACGCG-----CTGCGGCTGGGACCGCTGAAGTCGACATCGG 34093
Qy 1691 rSerGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuAr 1711
Db 34094 GCACGCGCGCGCTCGGGTGTGCGGTGTCAAGATGCTGCGCGCTGCGCGCA 34153
Qy 1711 gGlyLeuSerProArg-----GluSerSerLeuAlaLeuAsnTyrAlaAl 1726
Db 34154 CGGGGTGCTCCGAAGACCTCGACGTGGACGAGCGGACGACGAGTCTGCTGCGCG 34213
Qy 1726 aGlyPro----- 1728
Db 34214 CGGTTCGTGAGCTGCTCACCGGCGGTGGAGTGGCGGAGCGCGCGCGCGCGCGCGCG 34273
Qy 1729 ---ArgGlyIle-----IleAspLeuSerGl 1736
Db 34274 CCGGCGGCGCTCTCGCGTTCGCGGTGGCGGACGACGCGCGCTCTCTGAGGA 34333
Qy 1736 nValProHisLeu-----ProValLeuValProProThrProGlyThrProAlaTh 1753
Db 34334 GGCGCGCGGTTCGAGAGTCCCTGCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 34382
Qy 1753 rAlaMetAspAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSe 1773
Db 34383 -----GGCGTGGTCCCGTGGCGGTGTCGCGAAGACCTC 34417
Qy 1773 rSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSe 1793
Db 34418 GGCGCGCTGACCGCGCGCGAGTCCGG-----CAGTCGCGCGCAT 34456
Qy 1793 rSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSe 1813
```

```
Db 30819 GGCACACCCCGAGCT----- 30835
QY 530 luyysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLys-GluLys 549
Db 30836 -----CCCACTACGCTTCAGACCGAGCGCTTCGTGGTGAGAGC 30878
QY 550 ThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArg 569
Db 30879 TCCGCGCCACACGCGCGCGAC--GACTGGCGTTACCGCTGAGTGAAG---CCG 30932
QY 570 LysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGlu 589
Db 30933 CTACGCGCTTCGCGCCAGCGGACCTGTCCGGCGGTGGATCGTCGCGCGGAGCGAG 30992
QY 590 AlaAsnSerGluGluAlaIleThrProGlnInSerAlaGluLeuAlaSerMetGluLeu 609
Db 30993 CCAGAAGCCGAG-----CTGTGGCGCGCTAGAGCCGCGGAGCGGAGTTCGAGTA 31046
QY 610 AsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGlu 629
Db 31047 CTGGAAGCGGGCGGACGACGAGCGGTGAGGCCCTCGCGCCCGGCTCACCGCACTGACG 31106
QY 630 HisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCys 649
Db 31107 ACCGGCGACGCTTCACCGCGGTGTCTCGCTC----- 31139
QY 650 LysAsnPheTrpPheAsnTrpLysLysArgGlnAsnLeuAspGluIleLeuGlnHis 669
Db 31140 -----CTCAGCACCTCTGTGCCACAGGT- 31162
QY 670 LysLeu-----LysMetGluLysGluArgAsnAlaArgLysLysLysAlaPro 687
Db 31163 CGCTCGGTGACGACCTCGCGACGCGCGGAATCA-AGGCGCCCTGTGTGTCGTACCC 31221
QY 688 AlaAlaAlaSer---GluGluAlaAlaPheProProValValGluAspGluMetGlu 706
Db 31222 AGGCGCGGTCTCGTGGACGCTCGACACCCCGC----- 31257
QY 707 AlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAla 756
Db 31257 ----- 31257
QY 727 SerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSer 746
Db 31258 -----CCGACCCCGACCGGCGCATGCTCTGGGCGC-----TCGCGC 31293
QY 747 AspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGly 766
Db 31294 CGCTCGTGGCTTGAGCACCCCGAAGCGTGGCGCGGCTCTGTCGACC----- 31341
QY 767 ProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPro----- 782
Db 31342 -----TCCCGCGCCAGCCGATG-----CCGCGCGCTCGCCACCTCGTCACCG 31386
QY 783 -----ProThrProProArgArgThr-----SerArgAlaProIleGluProThrPro 798
Db 31387 CACTCTCCGCGCGCACCGGAGGAGGACGATCGCATCCGCACACCGGACTCCACGCCC 31446
QY 799 AlaSerGlu-----AlaThrGlyAlaProThr 807
Db 31447 GCGCGCTCGCGCGCACCCCTCCACGGAGCTGGCGCCACCCCGAGCTGGCAGCCCGCACG 31506
QY 808 ProPro-----ProAlaProProSerProSerAlaProProProValProLys 824
Db 31507 GCACGCTCTCATCACCGCGGACCGAGCCCTCGGACGCCACCGCGCAGCTGATGG 31566
QY 825 GluGluLysGluGluThrAlaAlaProProValGluGluGluGluGlnLys 844
Db 31567 CCCACACGCGGACGACCTCTCTCTGTCAGCGCGGAGCGGAGAACAGCCCGGAG 31626
QY 845 ProProAlaAla-----GluGluLeuAlaValAspThrGlyLysAla 858
Db 31626 -----GluGluLeuAlaValAspThrGlyLysAla 858
```


| | | | |
|----|-------|---|-------|
| Qy | 2075 | luGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro----- | 2088 |
| Db | 30507 | GC CGCCTGCGACGCTGCGCCAGCACCACCGCATCGGCACCTCTCTCGAAGCCATCCCGCC | 30566 |
| Qy | 2089 | -----LysGlnProGlyProValLysLeuGlyGlyGluAlaAla---- | 2101 |
| Db | 30567 | GAGACGCCCTCACCGCGTGTCTCAACCGCGCGCACCGCGGCGATCCGCTGGAC | 30626 |
| Qy | 2102 | -----HisLeuProHis----- | 2105 |
| Db | 30627 | GTCAACGGCCGGAGGACATCGCCGCATCTTGGCGCGAAGACGAGCGCGCGAGGTC | 30686 |
| Qy | 2106 | --LeuArgProLeuPro----GluSerGlnProSerSerSerProLeuLeu----- | 2120 |
| Db | 30687 | CTCGACGACCTGCTCGCGGCACATCGCGTTCGAGCGCTTCCTCTACTCTCTCGAAGCC | 30746 |
| Qy | 2121 | -----GlnThrAlaProGly--ValLysGlyHisGlnArgValValThrLeuA | 2136 |
| Db | 30747 | GGGGTCTGGGGCAGCGGACGAGGCGGTCTACGCGCGCGCCAAAGC----- | 30793 |
| Qy | 2136 | laGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeu | 2156 |
| Db | 30794 | -----CCACCTCGACGCGCTC | 30809 |
| Qy | 2156 | erAlaProLeuProAlaPro----- | 2162 |
| Db | 30810 | GCGCGCGCGCGCGCGCGCGGCGAGACCTCGTTCGCTCGCGGCGCTCTGGCGCC | 30869 |
| Qy | 2163 | --LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgProProSerA | 2182 |
| Db | 30870 | GCGACGGCATGGCGCGGGCGCGCAGCGGTACTGGCAGCTGCGCGCATCGCTCGG | 30929 |
| Qy | 2182 | spLeuTyrLeuProProAsp----- | 2189 |
| Db | 30930 | ATGAGCCCCGACGCGCCCTGGACGAACTGGCCNAGGCCCTGAGCCACGACGAGACTTC | 30989 |
| Qy | 2190 | -----HisGlyAlaProAlaA | 2195 |
| Db | 30990 | GTCCGCTGGCCGATGTCGACTGGGAGCGGTCGCGCCGCGCTTACGCTGTCCGCTCCC | 31049 |
| Qy | 2195 | rgGlySerProHisSerGluGly-----GlyLysArgSerProGluProAsnLysT | 2212 |
| Db | 31050 | AGCCTTCTGC-TCGACGGCGTCCCGAGGCCGCGCAGCGCTCGCGCACCGCTCGGTGC | 31108 |
| Qy | 2212 | hrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProGluGluGlyMet | 2232 |
| Db | 31109 | CCGGCTC-----CCGGCAGCGCGCG-----TGGC | 31135 |
| Qy | 2232 | hrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAsp----- | 2248 |
| Db | 31136 | GCCGACCGGCGAGTCTGTGGCGC-----TGGCGCGATCACCGCGCTCCCGAGCCCGA | 31189 |
| Qy | 2249 | -----GlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerG | 2267 |
| Db | 31190 | GCGCCGCGCGCGCTCTCACCTCGTTCGATCCACGCGCGCGCGCTACTCGGCCATT | 31249 |
| Qy | 2267 | lnProProAlaPhePheSerLysLeuThrGluSer---AsnSerAlaMetValLysSerL | 2286 |
| Db | 31250 | CTCCCCGACCGGGTGGCCCCCGCGCTCTCACCGAGCTCGGCTTCGACTCGCTCGAC | 31309 |
| Qy | 2286 | ysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnI | 2306 |
| Db | 31310 | GGCGTG-CAGCTCCGACACAGCTCTCCACCGTGTGTGGCAACAGG----- | 31355 |
| Qy | 2306 | leSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMet | 2326 |
| Db | 31356 | -----CTCCCGCGCACCGGTCTTCAGCACCCG-----ACGCCCGCGCACTCGCG | 31404 |
| Qy | 2326 | hrTyrArgSerGlnAlaVal-----GlnGluHisAlaSerThrAsnMetGlyLeuG | 2343 |
| Db | 31405 | CGCACTCCAGAGGGGTACTCTCGACCGCGCAGCGCGCCCGCAGGACTGG----- | 31458 |
| Qy | 2343 | luAlaIleAlaArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGluSerProProL | 2363 |

| | | | |
|----|-------|--|-------|
| Db | 31459 | AGGGGGGGTGGCCGGGGCCCTG-----GCCGAACGTGCC----- | 31493 |
| Qy | 2363 | euSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProI | 2383 |
| Db | 31494 | -----CTCGACCGGTGGGGACCGGGGGTCTCGACACCGTCTCTGC GCC | 31539 |
| Qy | 2383 | leThrAlaAlaAep-----GlyArgSerAspHisThrLeuThrSerProG | 2398 |
| Db | 31540 | TCACCGCATCGAGCCCGGCGGGTTCGGCGGGTTCGACGGCGCGCGCCGACCTG | 31599 |
| Qy | 2398 | lyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA | 2418 |
| Db | 31600 | GTGC-GGAGCC---GGAGCGGTGATCGACGACCTCGACGCCGA-----GGCCCTG | 31646 |
| Qy | 2418 | laProGlyLeuAlaSerGlyAspArgProPro---SerValSerValHisSerGluG | 2437 |
| Db | 31647 | ATCG-----GATGGCTCTGGCCCCCGTAACACTGACCGACCGCGGTCC----- | 31693 |
| Qy | 2437 | lyAspCys---AsnArgArgThrProLeuThrAsnArgValTrpGluAspArg----- | 2453 |
| Db | 31694 | ----TGCCCCACGCGCCGACCCCGCGCATCCCGCGCACCAACCGCCCCACACGCCA | 31748 |
| Qy | 2454 | -----ProSerSerAlaGlySerThrProPheProTyrAsnProLeu----- | 2467 |
| Db | 31749 | CAACCCCATCCAGACGCCGA---AGACCAACACCCAGATGACAGGTTCACAGCAAGTTG | 31805 |
| Qy | 2468 | -----lleMetArgLeuGlnAlaGlyValMetAlaSerPro-----ProProp | 2482 |
| Db | 31806 | GTGGACGCTCTGGCGCCCTCTCTCAAGGAGACGAGAACTCCGGNAGAGAGCGCTGC | 31855 |
| Qy | 2482 | roGlyLeuProAlaGlySerGlyProLeuAlaGlyProHis | 2495 |
| Db | 31866 | CGGGCGGACCGGTG-----GCAGGAGCCCAT | 31891 |

Db 28575 GAGCAGCGCGCTCGCGTCCGCGAGCAGCGGAGCGACTCGCGGACGCCCTCGCGATG 28634
Qy 1503 sTyrGluGluSerLeuLeuSerArgProGlyThrAlaSer-----SerSerGlyGl 1520
Db 28635 CCGGAGGAGCTGTGTACGCGCAGCTCTCGACGTCGGCGGCTGCGTCTCTCCCG 28694
Qy 1520 ySerIleAlaArgGlyAlaProValIleValProGluLeu----- 1533
Db 28695 GGCAGGCGCAGCAGTGGCGCGCATGGCGCGAACTCTTTCAGCAGCTCCCGGAGTTC 28754
Qy 1534 -----GlyLeuProArgGlnSerProLeuThrTyrGluAspHisGl 1547
Db 28755 GCTGCTCGATGCGCGAATCGGAGACCGCG-----TCTCCCGCTACGTCGACTGCTCTT 28811
Qy 1547 yAlaProPheAla-----GlyHisLeuProArgGlySerProValThrMetAr 1563
Db 28812 GAAGCCGCTGTCGAGCAGGAACCGCGGCAC-----CCACGCTGACCGCGCTCGACGTCGTC 28868
Qy 1563 gGluProThrProArgLeu-----GlnGluGlySerLeuSerSerSe 1577
Db 28869 CAGCCCGTGACTGCTGCTGTATGCTCGTGGCGAAGTCTGGCGACACCGCATC 28928
Qy 1577 r-----LysAlaSerGlnAspArgLeuLeuThrSerThrProArg---GluIleAl 1593
Db 28929 ACCCCAGCGCGCTGCTGCGCCACTCGCAGGCGAGATCGCGCGGTACGTCGCGGT 28988
Qy 1593 aLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHi 1613
Db 28989 GCACCTCACTCGACGACCGCGCGCTGCTCACCTCGCGCAGCAAGTCCATCGCGGCC 29048
Qy 1613 sLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAs 1633
Db 29049 CAC-CTGCGCGCAAGCGCGCATG-----ATCTCCCTCGCGCTCGA 29089
Qy 1633 pProThrSerIleProArgGlyIlePro-----LeuAspAlaAlaAlaTy 1649
Db 29090 CGAGCGCGCTGCTGAAGCGACTGAGCGACTTCGACGAGTCTCCGCGCGCGTCAA 29149
Qy 1649 rTyrLeuProArgHis-----LeuAlaProAsnPro-----ThrTyrProHi 1663
Db 29150 -CGGCGCCACCGCCAGCTGCTCTCGCGCGACCGCGACCGAGATCGAGAACTCGCCGCA 29208
Qy 1663 sLeuTyrProProTyrLeuIleArgGly-----TyrProAspTh 1676
Db 29209 CTG-----CGAGCGCGAGCGGCTCGGCGCGATCATCGGTCG 29250
Qy 1676 rAlaAlaLeu-GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMeth 1696
Db 29251 ACTACGCTCCACAGCGCGCAGTGTCTCGCGCGACCGCGACGATCGAGAACTCGCCGCA 29288
Qy 1696 sHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProA 1716
Db 29289 -----AAGGAGCTGGCGAGTCTCGCGGATCTCGCGGATCGCCCG- 29324
Qy 1716 rGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerG 1736
Db 29325 -----C 29325
Qy 1736 InValProHisLeuProValLeuValProProThr-----ProGlyThrProAlaThrA 1754
Db 29326 AGGCTCCGCGCAGCTGCGTCTTCT-----CCACCTCGAAGCGACCTGGATCACCGAGC- 29379
Qy 1754 lMetAspArgLeuAlaTyrLeuProThr-AlaProGlnProPheSerSerArgHisSer 1773
Db 29380 -----CGGTCGTCAGCGACCTACTGTACCGCAACTCGCCCATCGGTG-GCT 29430
Qy 1774 SerSerProLeuSerProGlyGlyProThrHisLeuThr-----LysProThrThrThrSer 1792
Db 29431 TCGCCCGCG-----CGGTGGAGACTTGGCGGTTCAGCGGCTTCACTTCATTCGAGG 29484
Qy 1793 SerSerGluArgGluArgAsp-ArgAspArgGluArgAspArgGluArgGluTy 1812
Db 29485 TCAGCGCCACCGCGCTCTCAACATGACCTCCCGAGACCGTCACTCGCGCTCGGACCC 29544
Qy 1812 sSerIleLeuThrThrThrThrThrValGluHisAlaProIleThrArgProGlyThrGl 1832
Db 29545 -----TCCGCGCGAACAGGAGCGCTGTGTCTCACTCACTCGCC 29591
Qy 1832 uGlnSerSerGlySerSerGlySerGlyGlyGlyGlySerSerArgProAl 1852
Db 29592 GAAGCTGGCGCAACCGCTCACCTACCTGGCGCGCATCTCCCGCACCGCAACCGCG 29651
Qy 1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGl 1872
Db 29652 CACC-----ACCCGAGCTCCCGACCTACG-----CC 29678
Qy 1872 nGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
Db 29679 TTCAGACCGAGCT-----TCTGGCTCAGAGCTCCCGCGCCACC 29720
Qy 1892 oSerLysProThrValLeuArgSerThrSerSerProValArgProAlaAlaTh 1912
Db 29721 AGCGCGCGCAGCAGCTGGCGTTACCGCTGAGTGAAGCGCTGACGCGCT- 29772
Qy 1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGly- 1927
Db 29773 -----CCGCGCAGCGGACCTGTCTCGCGCGTGTGTCTCGCGTGGGAGCGAGCAAA 29828
Qy 1927 ----- 1927
Db 29829 GCCGAGCTGCTGGCGCGCTGAAGCGCGGAGCGAGGTGAGTCTGTAAGAGCGCG 29888
Qy 1928 -----ValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPr 1943
Db 29889 GCGGACGACGACGCTGAGCGCTCG-----CGCGCGCTCACCGACTGACGAC 29939
Qy 1943 oArgValAlaArgProGluArgProArgAlaAspThr-----G1 1956
Db 29940 GCGGAGCTTTCACCGCGTGTCTCGCTCTCGACGACTCGTGCCACAGGTGCGCTG 29999
Qy 1956 yHisAlaPheLeuAlaLysProProAlaArg----- 1966
Db 30000 GTCAGGCACTCGCGCAGCGCGGAATCAAGCGCGCTGTGTCTCGTACCCAGGCGCG 30059
Qy 1967 -----SerGlyLeuGl 1970
Db 30060 GTCTCGCTGAGAGTCTTGACACCCCGCGACCCCGCGCGCGGCGGCGGCTC 30119
Qy 1970 uProAlaSerSerPro-----SerLysGlySerGluPr 1981
Db 30120 GCGCGGTCTGCGCTTGAGCACCCCGAACGCTGGCGCGGCTCGTCTGACCTCCCGCC 30179
Qy 1981 oArgProLeuValProPro-----ValSerGly-HisAlaThrIle-----AlaArgT 1997
Db 30180 CAGCGGATCGCGCGCTCGCGCCACTCGTCAACCGCACTCTCCGCGCGCACCGCGGAG 30239
Qy 1997 hrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPro-----A 2013
Db 30240 GACGAGATCGCCAT-----CCGACACCGGACTCCACCGCGCGCTCGCGCGCGCA 30293
Qy 2013 laProAlaSerAlaSerAspPro----- 2021
Db 30294 CCCTTCCAGGAGCTCGCGCCACCCCGAGCTGGCGAGCCCGCACCGCTCTCATCACC 30353
Qy 2022 -----HisArgGluTyThrGlnSerLysProPheSerIleGlnGluLeuLeuArgS 2040
Db 30354 GCGCGCACCGGAGCGCTCGCGCAGCCACG-----CCCA 30386
Qy 2040 erLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro-----ValS 2057
Db 30387 CGCTGGATGGCCACACGAGCGCGCAACCTCTCTCTCTCGTCAACCGCGCGGGAACAA 30446
Qy 2057 erProValSerSerProSer-----LeuThrHisAspLysGlyLeuProLysHisLeuG 2075
Db 30447 GCGCGCGGAGCGCACCACTCAACCGCAACTCAACCGCATCGGCGCGCGCTCACCATC 30506


```
QY 1924 rLeuAspGlyValTyPrThrLeuMetGluProValLeuLeuProLysGluAlaProAr 1944
Db 31893 -----CCGCGCGGTAGTCGCGCGGTGCCCTCACCTCGACGACGCGCT 31936
QY 1944 gValAlaArgProGlu-----ArgProArgAlaAspThrGlyHisAlaPheLeuAla 1962
Db 31937 CGTGTGTGACCTTCGACAGCAAGTCATCGCGCCACCTCGCGGCAAGGGGGCATG 31996
QY 1962 sProProAlaArgSerGlyLeuGluProAlaSerSerProSerLys----- 1977
Db 31997 ATCTCCCTCCCTCAGCAGGAAGCACCAGGAGCATCGAGAACTCCACGGACTG 32056
QY 1978 -GlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgTh 1997
Db 32057 TCGATCGCCCGCTCAACGGCGCTTACCGCCACCGTGTTCGGCGACCCACCCAGA-- 32114
QY 1997 rProAlaLysAsnLeu-----AlaProHisHisAlaSerProAs 2010
Db 32115 ----TCCAGAACTTGTCTAGCGGTGTGAGCGCGCATCCGCGACGGATCATCCCC 32170
QY 2010 pProProAlaProAlaSerAlaSerAsp-ProHisArgGluLeuLeuThrGlnSerLysP 2030
Db 32171 GTCGACTACGCTCCACAGCGCCAGTCGAGACCATCGAGAACTCGCGAGCTC 32230
QY 2030 roPheSerIleGlnGluLeuLeuArgSerLeuGlyTyHisGlySerSerTySerP 2050
Db 32231 CTGGCGGGTT----- 32241
QY 2050 roGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyL 2070
Db 32242 ----GTCCCCCAGACACCCCA---GGTCCCTTCTTCTCCAC--- 32277
QY 2070 euProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys- 2089
Db 32278 -----CTCGAAGGCACCTGGATCACCGAACCCCGCTCGAC 32314
QY 2090 -----GlnProGlyPro-----ValLysLeuGly---GlyGluA 2100
Db 32315 GCGCGGTACTGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCGCGCGTCGAGACC 32374
QY 2100 laAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuL 2120
Db 32375 CTGCCACCCAGCAGGCTTACCCACTTCATCGAGTCAAGCCGCCACCCCGCTCT--- 32430
QY 2120 euGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleS 2140
Db 32431 --CACCATGACCTCCCGCACAAGTCAACCG-----CCTGGCCACCC-T 32472
QY 2140 erGluValIleThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaProLeuP 2160
Db 32473 CCGACGCGAGACGCGGAGCAGCACCGCC---TCACCACCTCCCTTCGCGAGGCTGGGC 32529
QY 2160 roAlaProLeuTySerPheProGlyAlaSerCysProValLeuAspLeuArgArgProp 2180
Db 32530 CAACGGCTCGCCCTGACTGGGCTCCCTCTCTGCCCGCACGGGCGCCCTCAGCCCGC 32589
QY 2180 roSerAspLeuTyLeuPro-ProProAspHisGlyAlaProAlaArgGlySerPro--- 2198
Db 32590 CGTCC-----CCGACCTCCGACGTACGCTTCCAGCA-----CCGCTC 32628
QY 2199 -----HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeu 2215
Db 32629 GTACTGGATCAGCCCCCGGGTCCCGGAGGCGCGCGCACACCGCTTCGCGGCGCGA 32688
QY 2216 GlyGly-GlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGl 2235
Db 32689 GGCGCTCGCCGAGACGGGGCTCGCGTGGGCGCGGGTGGCGAGACCTCGACGAGGAGGG 32748
QY 2235 yHisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSe 2255
Db 32749 CCGG---CGCAGCGCCGTACTCGCGATGGTATCGGCA--GGCGGCTCCGTGCTCGGT 32804
QY 2255 rArgMetGlySerLysSerProGlyAenThrSerGlnProProAlaPheSerLysLe 2275
```

RESULT 85

```
AAZ87318
ID AAZ87318 standard; DNA; 36778 BP.
XX
AC AAZ87318;
XX
DT 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster.
```

| | | | | |
|---|---|-------|--|-------|
| D | B | 29839 | CCACGCGCGCACCCGCCGCATCCGGCGCACACCAGCCGCCGCCACACGCCCCACAAACCCCA | 29898 |
| Q | Y | 1387 | OProSerArgAspLeuThrGluAlaTyrIysThr-----GlnAlaLe | 1402 |
| D | B | 29899 | TCCACGAGCGGAAGACACACCCAGTAGACGAGTTCCAACGAACAGTTGTGTGAGCTCT | 29958 |
| Q | Y | 1402 | U-GlyProLeuLysLeuLysProAlaHisGluGluValAlaThrVal----- | 1418 |
| D | B | 29959 | GCGCGCTCTCTCAAGAGAGAACAAGAACTCCGGAAAGAGACCGTCGCCGGCGCAGCG | 30018 |
| Q | Y | 1419 | ----LysGluAlaGlyArgSerIle-HisGluIleProArgGluGluLeuArgHisThr | 1436 |
| D | B | 30019 | TCGGCAGGACCCCATCGCGATCTCCGATGACTCCGGTTCGGGGCGAATCCGGTC | 30078 |
| Q | Y | 1437 | ProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr--- | 1455 |
| D | B | 30079 | CCCCGAGGACCTCTGGGAGCCGCTCGCGCGGCGAAGACCTGGTCTCCGAGGTACCGGA | 30138 |
| Q | Y | 1456 | -----ProLeuLysTyrAspThrGlyValaSerThrThrGly | 1467 |
| D | B | 30139 | GGAGCGCGGTGGACATCGACTCCCTCTACGA-----CCCGTGCCCGGGCGCAAGGG | 30192 |
| Q | Y | 1468 | SerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProVal | 1487 |
| D | B | 30193 | -----CACGAGGTACGT-----CCGCAACCGCGCTTCTCTCG-ACG | 30227 |
| Q | Y | 1488 | HisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer | 1507 |
| D | B | 30228 | ACGCGCGCGATTTCGACGCGCTTCTTCGGAATCTCGCGCGAGAGGCCCTCGCCATCG | 30287 |
| Q | Y | 1508 | LeuLysSerArg-----ProGlyThrAlaSerSerSerGlyGlySer | 1521 |
| D | B | 30288 | ACCGCAGCAGCGCAGCTCTCGAACCTCTCGGAGGTCTTCGAGCGCGCGCATCG | 30347 |
| Q | Y | 1522 | IleAlaArg----- | 1524 |
| D | B | 30348 | ACCGCGCTCGGTCCGGCGACCGACGTCGCGCGTGTACGTGGCTGTGGTACACGAGACT | 30407 |
| Q | Y | 1525 | -----GlyAlaProValIleValProGluLeuGlyLys | 1535 |
| D | B | 30408 | ACGCGCGGACATCGGGTGCCTCCCGNAGGACCGCGGTACGTCTACCGGCAACT | 30467 |
| Q | Y | 1536 | Pro-----ArgLnsrProLeuThrTyrGluAspHisGlyAla | 1548 |
| D | B | 30468 | CCTCGCGGTGGCTCCGGCGCATCGGTACTCCCTCGACCTGGAGGACCCCGCGTGA | 30527 |
| Q | Y | 1549 | ProPheAlaGlyHisLeuPro--ArgLysSerProValThr----- | 1561 |
| D | B | 30528 | CCGTGGACAGCGGTGCTCTCTTCGTCTGCTCGCTCGCCTGCACCTCGCCCTGAAGGGCTCG | 30587 |
| Q | Y | 1562 | -----MetArgGluProThrProArgLeuGluGlu | 1571 |
| D | B | 30588 | GGACCGGACTGCTCGACGGCACTCGTGGCGCGGTGGCGTCTCTCGCAACCGCGGGCG | 30647 |
| Q | Y | 1572 | GlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlu | 1591 |
| D | B | 30648 | CGTTCA---TCAGTTACGACGACGAGGACCATGGCCGCGACGCGGACCAAG-- | 30701 |
| Q | Y | 1592 | IleAlaLysSerPro-----HisSerThrValProGluHisHisProHisProIleSer | 1609 |
| D | B | 30702 | -----GCTTCGCTCGCGGGCGGACGCTCGCTCGCTGGGGCGAGGGGTCTCGCGTACTCC | 30755 |
| Q | Y | 1610 | -ProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrrargSerHisIlePr | 1629 |
| D | B | 30756 | TCCTCGAACCGGTCTCCGACGCGCGCGCAAGG----- | 30789 |
| Q | Y | 1629 | oLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaTaTy | 1649 |
| D | B | 30790 | -----CCACCGGT---CCTGGCGCTCTGTGCGCGCGACG | 30821 |
| Q | Y | 1649 | rTyrLeuProArgHisLeuAlaProAsnPro-ThrTyrProHisLeuTyrProTyrL | 1669 |
| D | B | 30822 | CCATCAACGAGGACGCGCGACCAACCGGCTCACGCTCCGAC----- | 30865 |

| | | | |
|----|-------|--|----------|
| QY | 864 | erGluCysThrGluGluAlaGluGluGlyProAlaLys | -----876 |
| Db | 27790 | CGGCCACACCCCGAGCTCCACCTACGCTTCCAGACGAGCGCTTCTGGCTGCAGAG | 27849 |
| QY | 877 | -----GlyLysAspAlaGluA | 882 |
| Db | 27850 | CTCCGGCCACACGAGCCCGCCGACGACTGCGCTTACCGGCTCGAGTGGAAAGCCGCTGAC | 27909 |
| QY | 883 | -----AlaGluAlaThrAlaGluGluAlaLeuLysAlaGluLysGluGlyGlys | 900 |
| Db | 27910 | GGCCTCCGGCCAGCGGACCTGTCGGGC | 27945 |
| QY | 900 | er | 900 |
| Db | 27946 | CGTCGCGCTCGGAGCGAGCCAGAAGCCGAGCTGCTGGGCGCGCTGAAGGCCGCGGAGC | 28005 |
| QY | 901 | -----GlyArgAlaThrThrAlaLysSerSer | 911 |
| Db | 28006 | GGAGGTGACGTACTGGAAGCCGGGGCGGACGACCGGTGAGGCGCTCGCCCGCGCT | 28065 |
| QY | 911 | laProGln | 921 |
| Db | 28066 | CACCGCACTGACACCGCGGACCGCTTACCAGCGGTGCTCTGCTCTGACGACCTCGT | 28125 |
| QY | 921 | ys | 933 |
| Db | 28126 | GCCACAGGTGCGCTGGGTGAGGCACTCGCGGACCGCGGAATCAAGCGCGCCCTGTGGTC | 28185 |
| QY | 933 | sPlysAsnArgLeuLeuSerProArg | 948 |
| Db | 28186 | CGTACCCAGCGCGCGGTCTCCGTCGGACGTCTCGACACCCCGCGC | 28239 |
| QY | 948 | hrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysG | 968 |
| Db | 28240 | CCGGGCATGCTCTGGGGCTCGCGCGCTGCTGCGCCTTGACGACCCCGAAGCTGGGC | 28299 |
| QY | 968 | lnArgAlaAlaIleProProIlelnValThrLysValHisGluProProArgGluA | 988 |
| Db | 28300 | CGGCTCTGTCGACCTCCCGCCGAGCCGATG | 28347 |
| QY | 988 | sPalaAlaProThrLysProAlaProAla | 1003 |
| Db | 28348 | CTCTGTCACCGCACTCTCCGGCGCCACCGCGAGGACCATCGCCATCGCACCCCGG | 28407 |
| QY | 1003 | lnAsnLeu | 1010 |
| Db | 28408 | ACTCCAGCCCGCGCTCGCCCGCGCACCCCTCCAGGAGCTCGGCCACCCGCGACTG | 28467 |
| QY | 1010 | sPalaProGlnProGlySerProArgGlyLysSerArgSerProAlaProPro | 1029 |
| Db | 28468 | GCAGCCCAACGACCGCTCTCATCACCGCGGACCGGAGCCCTCGGACGACCGCGC | 28527 |
| QY | 1030 | -----AlaAspLysGluAlaPheAlaAlaGluAlaGlnLysL | 1042 |
| Db | 28528 | ACGCTGGATGGCCACACGAGCCGAACACCTCTCTGTCGACGCGGAGCGGCGGAACA | 28587 |
| QY | 1042 | euProGlyAspProProCysTriThrSerGlyLeuProPheProValProProArgGluV | 1062 |
| Db | 28588 | AGCCCCGGAGCCACC | 28635 |
| QY | 1062 | alileLysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaProProGlyH | 1082 |
| Db | 28636 | CCGCGTCACCATCCGCGCTGCGACGTGCGCGGACCCCGCCATGCGCACCTCTCTCGA | 28695 |
| QY | 1082 | isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrI | 1102 |
| Db | 28696 | CGCCATCCCG | 28727 |
| QY | 1102 | leSerAsnProProProLeuIleSerSerAla | 1113 |
| Db | 28728 | --TCGTCCACACCGCGCGCACCGGGCGCGATCGCTGGAGCTCACCGGCCCGGAGGA | 28785 |
| QY | 1113 | ysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnL | 1133 |
| Db | 28786 | CATCGCCCGCATCTTGGCGCGAAGACGAGCGCGCGAGTCTCTCGACGACCTGTCTCG | 28845 |
| QY | 1133 | euHisValProTySerGluHisAlaLysAlaProValGlyPro | 1147 |
| Db | 28846 | CGGCACTCGCTGGAGCCCTTGTCTCTACTCTCTGAACGCCGGGTCTGGGGGACGCG | 28905 |
| QY | 1148 | -----ValThrMetGlyLeuProLeuPro | 1162 |
| Db | 28906 | CAGCCAGGCGCTTACGCGCGCGCAACGCCACCTCGAGCGCTCGCCCGCGCGCG | 28965 |
| QY | 1162 | laProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyProProG | 1182 |
| Db | 28966 | CGCCC | 29016 |
| QY | 1182 | luser | 1195 |
| Db | 29017 | CGGATGGCGCGGCGCGGACGCGCTACTGGGAGCGTTCGGGATCGCTCCGATGAG | 29076 |
| QY | 1195 | ----- | 1195 |
| Db | 29077 | CCCCGACCGCCCTGGACGAAGTGGCAAGCCCTGAGCCACGACGACCTTCGTGCG | 29136 |
| QY | 1196 | -----ArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys | 1210 |
| Db | 29137 | CGTGGCGGATGCTGACTGGGAGCGGTTCGCG | 29195 |
| QY | 1211 | -----GlyIleProSerThrArg | 1217 |
| Db | 29196 | TTCGTGTCGACGCGCTCCCGAGGCGCGGAGCGCTCGCGCACCCGCTGTCGCGCCG | 29255 |
| QY | 1217 | alProSerAspSerAlaIleThrTyArgGly | 1230 |
| Db | 29256 | CTCCCGCGACCGCGCGTGGCGCGCACCGGCGAGTCTGCGCGCTGCGCGCATCACCG | 29315 |
| QY | 1231 | -----HisGlyThrProAlaAspVal | 1245 |
| Db | 29316 | CGTCCCCGAGCCCGAGCGCGCGCGCTCTCACCTCTGCTCCGACCGCGCGG | 29375 |
| QY | 1245 | rgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProL | 1265 |
| Db | 29376 | CGTACTCGGCATCTCTCCCGACCGGTGCGCCCGCGCTGCTCCAGCGTGGTGGC | 29491 |
| QY | 1265 | ysGlyHisValIleTyGluGlyLysGlyHisValLeuSerTyGluGlyGlyMets | 1285 |
| Db | 29433 | TCGGCTTCA-CTCGCTGACGCGCGTGACCTCCGCAACGAGTCTCCAGCGTGGTGGC | 29491 |
| QY | 1285 | erValThrGlnCysSerLysGluAspGly | 1298 |
| Db | 29492 | AA-----CAGGCTCCCGCCACCGCTTTCGACACCCGACCGCGCGCGACTCGCC | 29545 |
| QY | 1298 | erGlyProProHis | 1314 |
| Db | 29546 | GGGCACTCCAGAGCGTACTCGCACCGCGCGGAGCGCGCGCGCGCG | 29599 |
| QY | 1315 | GluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAla | 1334 |
| Db | 29600 | GAGGGCGGGTGGCGCGGCGCTGGCGGAACTGCCCCCTCGACCGGTGGG-GGACGGGG | 29658 |
| QY | 1335 | IleProProGluArgHisSerProHisHisLeuLysGlnHisIleArg-Glyse | 1354 |
| Db | 29659 | GGTCTCGACACCGCTCTCGCTCACCGGATCGAGCGCGCGGTTCGCGCGGTTTC | 29718 |
| QY | 1354 | rIleThrGlnGlyIle | 1373 |
| Db | 29719 | GGAGCGCGCGCGCGACCTCGTGGTGGAGCGCGGAGCGGTGATCGACGACCTGGACGC | 29778 |
| QY | 1373 | gGluAlaLysLeuLeuLysArgGluGly | 1382 |
| Db | 29779 | CGAGGCGCTGATCGGATGGCTCTCGCCCCCGGTAAACCTGACCGCGCGGTCTCTGC | 29838 |
| QY | 1383 | -----ThrProProProPr | 1387 |

Db 25858 ---GGACCTTTCTGTCAGTTTCAG-----CCGCGAGCGCGGCTGGCCGCGGACGG--- 25905
Qy 330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgLysGlnArgLysGlnGluArg 349
Db 25906 -----CCGGTGAAGAGGCTTTCGCACCTCGCGGACGG 25938
Qy 350 MetGlnSerArgValGlyGlnArg-----GlySerGlyLeuSerMet-SerAlaAlaAr 367
Db 25939 CTTTCGGCCCGCGAGGCGCGGAGTCTGCTGCTGGAGCGCTCTGTCGACCGCCGCGC 25998
Qy 367 gSerGluHisGluValSerGluIleLeu-----AspGlyLeuSe 380
Db 25999 CAACGGACACCGGATCTCCGGGTGCTCCGGGAGCGCGGTCAACACGAGACGGCGCCAG 26058
Qy 380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db 26059 CAACGGCTCAGCGCTCCGACGGCGCTCCAGCAGCGGTCTATCCGACGGCGCTTGGC 26118
Qy 400 rAspAla-----AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLe 415
Db 26119 GGACGGCTCGCGCGGTGAGTGGAGCTGTCGAGCGCGCGACGGCACGGCGACGGC 26178
Qy 415 uMetAlaAspProMetLys-----ValTyrLysAsp-ArgGlnValM 429
Db 26179 GCTCGGCGACCGGATCAGCGCGAGCGGCTCATCGCCACCTACGGCCAGGAGAGCAG 26238
Qy 429 etAsnMet-----TrpSerGluGlnGluLysGlnThrPheArgGluLysPheMetG 446
Db 26239 CGAACACGGCTGAGGCTGGCGGCTTGAAGT-----CGAACATCGCGGCACAC 26286
Qy 446 lnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluC 466
Db 26287 CGACGGCGCGCGCTGTCGAGGTGTCATCAAGATGCTCCAGCGCGATGCCACGGACT 26346
Qy 466 ysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgS 486
Db 26347 GCTGC----- 26351
Qy 486 erTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 506
Db 26352 -----CGAAGACGCTCAGCGTCGACGAGCGCTCGAGCCAGATCGACTGGTCGGCGGCAC 26406
Qy 506 lnGlnGlnGlnProMetProArgSerSerGlnGluLysAspGluLysGluLysG 526
Db 26407 GGTGAAGTCTCACCAGGCGCTGCTGCTGCGGCG-GAGAAGCAGCAGCGCGGCTGCGCC 26465
Qy 526 lu-----LysGluA 529
Db 26466 GCGCGGCTGTCTCTCTTCGTCATCAGCGGAGCGAACGCGCAGCTGCTCTGGAGGAG 26525
Qy 529 laGluLysGluGluLysProGluValGlu----- 539
Db 26526 CCGCGCGCTCGAGGACTCCCGCGCGCTCGAGCGCGCGCGCTGCTGCTGCTGCTGCTG 26585
Qy 539 ----- 539
Db 26586 GCGCGGTCTCGCGAGACTCGCGCGCTGAGCGCCCGACATCGGCGAGCTCGCGCGT 26645
Qy 540 -----AsnAspLysGluAspLeuLeuLysGluLysGluLysThrAspAspThrSerG 555
Db 26646 ACGCGGACGCTGTCAGGACGTGGATCCGGGCTGGCGCGCGCGCGCTGCTGCTGCTGCTG 26705
Qy 555 lyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerG 575
Db 26706 GTACGGCGATGAGCAGCGCGCTGCGGTCGGCGCAGCGCGGAGGCACTCGCGGAGC 26765
Qy 575 lnglyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluA 595
Db 26766 CCTCGGATGCGGAGGACTGTGTACCGGCGACGTCTCGAGCTGGCGCGCGGCGGCTG 26825
Qy 595 laileThrProGlnGlnSerAlaGluLeuAlaSerMet-----GluLeuAsnGluSerS 613
Db 26826 TCGTCTTCCCGCGCAGGCGCAGCGCAGTGGCGCGGATGGCGCGGCGGCGGCGGCGGCGG 26885

Qy 613 er-----ArgThrPThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluH 630
Db 26886 CACCGAGTTCTGCTGCTGATGGCGGATCGAGACCGCG-----CTCTCCC 26933
Qy 630 isGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysL 650
Db 26934 GCTACGTCAGCTGGTCT----- 26950
Qy 650 ysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisL 670
Db 26951 -----CTTGAAGCCGCTGCTCCGACAG----- 26971
Qy 670 ysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaA 690
Db 26972 -----GAACC-CGGCGCACCCACGCTCGACCGGTGACGTCTCCAGCC----- 27015
Qy 690 laSerGluGluAlaAlaPheProValValGluAspGluGluMetGluAlaSerGlyV 710
Db 27016 -----CGTGACCTTCGCTGTCATGGTC 27037
Qy 710 alSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnG 730
Db 27038 TCGCTGCG-----GAAGGTCTGGCA-----GCACCAGGCATCACCCCGAGCGCGTC 27085
Qy 730 luValPro-----ArgGlyGluCysSerGlyProAlaAr 741
Db 27086 GTCGGGCACTCGCAGCGGAGATCGCGCGGTGCTGTCGCGGTGCACTCA-CCCTCGA 27144
Qy 741 hrVal---AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaL 760
Db 27145 CGACGGCGCGCGGTGCTGCTCACCTGCGCAGCAAGTCCATCGCGCGCCACCTCGCGCGG 27204
Qy 760 ys-----AspThrG 763
Db 27205 GGCGGCATGATCTCCTCGCCCTCGACGAGGCGCGCTCTCTGAAGGACTGAGCGACTT 27264
Qy 763 lyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProGlyProp 783
Db 27265 CGACGGACTCTCGTCGCGCGCTCAACG-----GCCCGACCGCCACCGCTGCTCCGG 27318
Qy 783 roThrProProArg-----ArgThrSerArgAlaProIleGluProThr----- 797
Db 27319 CGACCGGACCCAGATCGAGGAACCTCGCCGACCTCGAGGCGCGACGGCTCGCGCGCG 27378
Qy 797 ----- 797
Db 27379 GATCATCCCGTCTGACTAGCTCCACACGCGCGAGGTCTGAGTTCATCGAGAAGGAGCT 27438
Qy 798 -----ProAlaS 800
Db 27439 GGCGGAGTCTCGCGGACTCGCCCGCGAGGCTCGCGACGTGCTGCTTCTTCTCCACCT 27498
Qy 800 erGluAlaThrGlyAlaPro----- 806
Db 27499 CGAAGGCACCTGATCACCGACCGCGGTCTCGACGGCACCTACTGTGTACCGAACCTGCG 27558
Qy 807 -----ThrProProAlaProPro----- 813
Db 27559 CCATCGGTGGTCTCGCCCGCGCGGTGGAGACTTGGCGGTGAGCGCTTACCCACTT 27618
Qy 814 --SerProSerAlaProProValValProLysGluGluLysGluGluGluThrAlaA 833
Db 27619 CATCGAGTTCAGCGCCCGCCCGCTCTCTCAGCATGACCTCCCGAGAGACCGTCAACCGCT 27678
Qy 833 laAlaProValGluGluGlyGlu-----GluGlnLysP 845
Db 27679 CGGACCCCTCGCGCGCGAAGCAGGGAGCGGAGGCTGCTGCTACCTCACTCGCGGAGC 27738
Qy 845 ro-----ProAlaAlaGluGluAlaValAspThrGlyLysAlaGluGluProValLysS 864
Db 27739 CTGGGCCAACGGCC-----TCACCATGACTGGGCGCGCCATCTCTCCCGACCGCAAC 27789

Db 23000 TCAGCCTACCCCGAGCCCCCTCTGCTGTCTCGC-----TGTATCTCCTC 23044

QY 2479 oProProGlyLeuProAlaGly---SerGlyProLeuAlaGlyProHisAla-T 2498

Db 23045 TGCACAGTTGGGTCTCATGCCGGTGTTCAGGCCCT-----CCTTGGCACAGAT 23095

QY 2498 rPaspGluLeuProLysProLeuLeuCysSerGlnTyr 2510

Db 23096 GGGCCACAGATACTGTCCCGCGCTCTCCCTGCACATAT 23133

RESULT 84

ABS56090

ID ABS56090 standard; DNA; 38506 BP.

AC ABS56090;

XX

XX 21-JAN-2003 (first entry)

XX

DE S. venezuelae DNA inserted into cosmid pKOS023-27.

XX

XX Narbonolide polyketide synthase; PKS; desosamine biosynthetic gene;

KW desosaminyl transferase gene; beta-glucosidase gene; antibiotic;

KW pick hydroxylase gene; C12 hydroxylase gene; narbonolide;

KW desosaminylated polyketide; narbomycin biosynthesis; mutant;

KW picromycin biosynthesis; ds.

XX

OS Streptomyces venezuelae.

OS Synthetic.

XX

XX WO200297062-A2.

XX

PD 05-DEC-2002.

XX

PF 22-FEB-2002; 2002WO-US005642.

XX

PR 22-FEB-2001; 2001US-00793708.

XX

PA (KOSA-) KOSAN BIOSCIENCES INC.

XX

XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;

XX WPI; 2003-041412/03.

XX

PT Preparation of polyketides by recombinant DNA technology, useful as

PT antibiotics and as intermediates in the synthesis of pharmaceutical

PT compounds.

XX

PS Disclosure; Page 20-30; 127pp; English.

XX

CC The present invention relates to recombinant DNA sequences encoding for a

CC narbonolide polyketide synthase (PKS) domain, and methods of producing

CC polyketides by recombinant DNA technology. The recombinant DNA sequences

CC are derived from Streptomyces venezuelae desosamine biosynthetic.

CC desosaminyl transferase, beta-glucosidase, or pick (C12) hydroxylase

CC genes. The method is useful for transforming a cell with a recombinant

CC expression vector that encodes a functional beta-glucosidase gene, and

CC therefore for increasing the yield of a desosaminylated polyketide in a

CC cell. The recombinant methods and materials are useful for expressing

CC polyketides with significant antibiotic activity, derived in whole or in

CC part from the narbonolide PKS gene, and other genes involved in

CC narbomycin and picromycin biosynthesis in recombinant host cells. The

CC present sequence represents S. venezuelae DNA inserted into cosmid

CC pKOS023-27 in the methods of the present invention

XX

SQ Sequence 38506 BP; 4914 A; 15116 C; 13446 G; 5030 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.61e-09 Length: 38506

Score: 601.00 Matches: 700

Percent Similarity: 31.77% Conservative: 322

Best Local Similarity: 21.76% Mismatches: 1144

Query Match: 4.55% Indels: 1067

DB: 7 Gaps: 148

US-09-522-753-5 (1-2517) x ABS56090 (1-38506)

QY 53 LeuSerProGlySerLeuLeuGlnProGlnArgProSerLeuLeuSerGluPhe 72

Db 24888 CTGCGCCCGCCCGCGCGCGCTCGAGCT--CGCCCGCGCGCTAGGGCGCTGGCCGC 24946

QY 73 GlnProGlyAsnGluArgSerGlnGlu-----LeuHisLeuArgPro 86

Db 24947 GGCCTGGGGACGACGGCGACGCCACCGACCTGGAGCGCGTCCGACGACCT 25006

QY 87 GluSerHisSerTyrLeuProGluLeuGlySerGluMetGluPheLeuSerLys 106

Db 25007 -----CTTCTCTTCATCGACAAGAGCTGGCGACTCCGACTTCTGACCTCCCG 25057

QY 107 ArgProArgLeuGluLeuProAspProLeuLeuArgProSerProLeu----- 123

Db 25058 ACACCA-----CCGGCACCAACCGCGCACCCAGCCCGCCCTCACACCGGA 25102

QY 124 -----LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138

Db 25103 ACACGGAACGACAGCGGAGACCGGAGCCATGGCGAACAACGAGACAAGCTCCGGAC 25162

QY 139 -----ArgSerLeuThr 142

Db 25163 TACCTCAAGCGCTCACCGCGAGCTGCAGACAACACCGAGCGTCTCGCGAGATCGAG 25222

QY 143 GlyLysLeu---GluProValSer-----Pro 150

Db 25223 GGACGACGACGACCGCGTGGCGATCGTGGGATCGCGCTCGCGCGGGGTGTC 25282

QY 151 ProSerProProHisThr-----AspProGluLeuGluLeu---Val 163

Db 25283 GCCTCGCCCGGAGACCTGTGGCAGCTGTGGCGGAGCGGCGGACGC-GATCTCGAGTT 25341

QY 164 Pro-----ProArgLeuSerLysGluLeuLeuGln---AsnMetAspArgValAsp 180

Db 25342 CCGCAGGACCGCGCTGGAGCGTGGAGGGGCTGTACAGCCCGCCCGACCGCGCTCCGG 25401

QY 181 ArgGluLeuThrMetValGluGlnGlnLeuSerLysLeuLysLys---LysGlnGln 199

Db 25402 CAGGACGTACTGCGCGTCCGGCGGATTCCTGCACGACGCGCGGAGTTTCACGCCGACTT 25461

QY 200 LeuGluGluAlaAlaLysProPro----- 208

Db 25462 CTTGCGGATCTCGCGCGGAGGCGCTCGCCATGAGCCGACGACGAGCTGTCCCTCAC 25521

QY 209 -----GluProGluLysProValSerProPro 217

Db 25522 CACCGCTGGGAGCGATCGAGAGCGCGGCGCATCGACCGCGCCCTGAAGGCGACGCG 25581

QY 218 ProIleGluSerLysHisArgSerLeuValGlnIleTyrAspGluAsnArgLysLys 237

Db 25582 CCTCGCGCTTCTGTCGGCGGTGGCACACCGGCTACACCTCGGGGACGACACCGCGCT 25641

QY 238 AlaGluAla-----AlaHisArgIleLeuGluGlyLeu-----Gly 249

Db 25642 GCAGTCGCGGAGCTGAGGGCCACCTGTGTGTCAGCGCGCGCGCTGCTGTCTCCGG 25701

QY 250 ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsn 269

Db 25702 CGGTATCGGTACGTCTCGGTACGAGCGGACCGCGCCCTGACCGTGGACGCGCTGCTC 25761

QY 270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289

Db 25762 GTCTCGTGTGCGCCCTGCACCTCGCGTGGAGGCCCTCCGCAAGGCGAGTGGCGCAT 25821

QY 290 AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309

Db 25822 GGCCCTCGCGGTGGTGTACGGTGTATGCCAACGC----- 25857

QY 310 GluLysLysValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysVal 329

DB: |||||

QY 547 LysGluLysThrAspSerThrSerGlyGluAsp-----AsnAspGluLysGluAlaVal 564
Db 17724 CTTGAAAAGTTGAAGAAGAGGCTCTCCATGATACAGAGAAGACTCCAACGATAGATC 17783
QY 565 AlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgLysGly----- 580
Db 17784 AGCAGGTCCTCTGACACTTCACGACCGAGGAGGAGCGGGGTGACTGCCTTC 17843
QY 581 -----ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThr 597
Db 17844 CTGGAGTCTCTTTCAGGAGCCACCTGGAGACGACGAAGCGGAGGAGCGCTGCCTC 17903
QY 598 ProGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTrpThrGlu 617
Db 17904 CCTGAGAAGCTGAAGAAGAGGAGGACGACAGACTCTCTCATCTTCATCCAAAGAAGAGC 17963
QY 618 GluGluMetGluThrAlaLysLysGlyLeuLeuGlu-HisGlyArgAsnTrpSerAlaIle 637
Db 17964 CACGCCGAGGAGCGACCAAGAAGAGAGCGCGAGAGAAAGAGAGGCGGCGAT 18023
QY 637 eAlaArgMetValGlySerLysThrValSerGlnCysLys----- 650
Db 18024 AAGGAGGCGGTAGCAGGAAGACTCCGCGCAGTACGAAAGAGGACTTCTTGAGGCGGAT 18083
QY 651 -----AsnPheTrpPheAsnTrpLys-----ArgGlnAsnLeuAspGluIleLe 666
Db 18084 GCTTACGAGGTTCT-TACAAATGAAGCTGACATAGAAGATGAGCTAGATAAACCAT 18142
QY 666 uGln-----GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgLys 682
Db 18143 TGAATTGTTTCTACCGAAGAAGATAAATGATTCCGAGAGAGAACTTCCAAGAA 18202
QY 682 sLysLysLysAlaProAlaAlaSerGluGluAlaAlaPheProValValGluAs 702
Db 18203 AATGAAAGAGAACTAAAGCCTTATGATCTAGTGCCATCAA-CATCTTAAAGAGAAGA 18261
QY 702 pGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaG 722
Db 18262 AGAAGAGAGAGAAACACAGGGAGAAATGGAGAGACGAGAA----- 18301
QY 722 uAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVa 742
Db 18302 -----CGAGAGGACCGCGGACGAGCATCGGATGGGTGCTGCGGC 18342
QY 742 lAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaLysAspTh 762
Db 18343 A-----TCACAGGACGAGCTCTCGCGCATCACAGGACGAGCAGAGCCGCCACCA 18396
QY 762 rGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyPr 782
Db 18397 GGGACAGGACAGCCCGCGCTGCTCAAGACAGTCCAGGAGGAGGAGGCGCGA-GG 18455
QY 782 oProThrProPro-----ArgArgThrSerArgAlaProIleGluPro-----Th 797
Db 18456 CTCGCGGATGCCMAACTCAAGAGAGAAATCAAGACGCT-CCAGAGAAAGAAAGGCGCA 18514
QY 797 rProAlaSerGluAlaThrGly-----AlaProThrProPro----- 810
Db 18515 CCCACTGAAGATGAGCAACGGAATGATAGTACGCCATCCAAAGACCCAGGCAAGAA 18574
QY 811 ----AlaProProSerProAlaProProValProValProLysGluLysGluG 829
Db 18575 AGACCCAGGCGCCAGGGA-GAAGCTCTCTGGGAGCGGACCTGATGATGATGATGATG 18633
QY 829 uGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaG 849
Db 18634 AGAGGATGCTGTC-----CCAGAGGACCTGGAGATCGAGAGCGCCACCAAGCGGCA 18687
QY 849 uGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluG 869
Db 18688 AGGAGGATGAGCAATGAGAGAGCTGAGGACCGTCCG----- 18730
QY 869 uAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaThrAlaGluG 889

Db 18731 -----AGACCCCAAGCTCAAGGAGAAGG 18753
QY 889 yAlaLeuLysAlaGluLysLysGluGlySerGlyArgAlaThrThrAlaLysSerSe 909
Db 18754 CGAAGCCGCGACAGCAGCGGCGAAGAGGCTCTGCAATTCCTGCTAAGAAACCCCGG 18813
QY 909 rGlyAlaPro-----GlnAspSerAspSerSerAlaThrCys-S 922
Db 18814 GCTCGACCCCTCCATTTAAAGACAAAAGCTCAAGAGTGCATCTTATCCACCTGCGG 18873
QY 922 eAlaAspGluValAspGluAlaGluGlyAspLysAsnArgLeuLeu----- 938
Db 18874 CGGAAATAAGCTACACCCAGCATCAGGTGCAGACTCCAAAGACTGGTGGCAGGCCCTC 18933
QY 939 -----SerProArgProSerLeuLeuThrProThrGlyAspP 951
Db 18934 ACATGAAAGAGTCTCGCTCCCGCAGGCCCTGACAGAGCCGCGCCACTGCGCTGC 18993
QY 951 rArgAlaAsnAlaSerProGlnLysProLeuAspLysGlnLeuLys-----GlnArg 970
Db 18994 CC-----ACCTTACGTGGTGCTATC-CTGCCCCAGCTACGAGGAGGTGATG 19040
QY 970 laAlaAlaIleProIleGlnValThrLysValHisGluProProArgGluAspAlaA 990
Db 19041 CACAGCCCGAGGAGCCCGTCTCTGCAGCGCGCTAGTACGCGGACTCTGTGT-TGACTG 19099
QY 990 laProThrLys-----ProAlaProAlaProProProPro- 1002
Db 19100 CGCGGACTCGCAGACTCCACGCCCGTGCACCGCTCCACACAGCGCTGTCTCCCTC 19159
QY 1002 ----- 1002
Db 19160 CTTTTTCGACAGGTTCTCCGTGCTTCAAGTGGGCTTCGGAACCGCAGCGAGCTCC 19219
QY 1003 --GlnAsnLeuGlnProGluSerAspAlaProGlnGln-----ProGlySerSerP 1019
Db 19220 TGCAGGCTCTCTCCACAAACCTTTACCGCTCGGTCTCTGTGCGACATTAGGAGGACCC 19279
QY 1019 rArgGlyLysSerArgSerProAlaProAlaAspLysGluAlaPhe---AlaAlaG 1038
Db 19280 CGAGGAAGAAATTCAGCTCGAGACAAAGCTTTCAGGACGAGAGGCTTCTGCTGCTCCTC 19339
QY 1038 luAlaGlnLysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProValP 1058
Db 19340 CAGCTACGACTCTCCATGCCACCTCGA---TGAAGACAGGCGCCCTGCCCCCGGT 19396
QY 1058 roProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrA 1078
Db 19397 TCCCGCGAGAAAGTTTGGCTGCTTGTGCGCAGGGTACTACTCCACAGACTATGCGCTCCC 19456
QY 1078 laProPro-----GlyHisProLeuProLeuGlyLeu-HisAspThr 1091
Db 19457 GTCCGCCAAGTCCAGCTTTGACCTGCCACCGCTGCGGTGTCTGCTGCTGCTGCTGCT 19516
QY 1092 AlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSer 1111
Db 19517 TCCAGAGGGCTCTTCTCAAG-----TTTACAAGC 19546
QY 1112 AlaLysHisProSerValLeuGluArgGlnIleGlyAlaIle---SerGlnGlyMetSer 1130
Db 19547 AAAACCTTCCCTTCCCGCAGCGCGAGCTGCTGCTGCTTCTTCTTCTCGAAGGGCGCTTCC 19606
QY 1131 ValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly-ProValThrMe 1150
Db 19607 -----CCCGACCTGGACACTCCGAGGACAGCAGGAGCGCGCGCAT 19651
QY 1150 tGlyLeuProLeuPro-----MetAspProLysLysLeuAlaProPheSerGlyVal 1168
Db 19652 CATCCCCCGAGGCCAGCTACCTGGAGCGCTGGACGAGGGTCCCTTCAGCGCGCTCAT 19711
QY 1168 sGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProTh 1188

Db 5162 GCCTGCTCCACGAGCCCTCGAGACATGCCTACGCC---GAGGCGCTCAAGCCCTCA 5106
Qy 2057 erProValSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGlu 2077
Db 5105 AGCGCTTCGCTCCCTCTCC---CACCTCTCTCC 5073
Qy 2077 euAspLysHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys--- 2095
Db 5072 TGGGAGCGGACACCTCTCTGACGACCT-CGTCAAAGAACCGCAACCTGGCGACACC 5014
Qy 2096 -----LeuGlyG 2098
Db 5013 CTCAACGCTGCATGCCCTTCTACCCCGGCTGGCACTCGCGCTCCCTCAACTACAG 4954
Qy 2098 lyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerP 2118
Db 4953 GTCAGCGCTGATCGCTGTCATCTAGGTCACCGCTCCCGCAAGTCCCGAGCTGCTC 4894
Qy 2118 ro-----LeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrL 2135
Db 4893 CGCAACCTCATCTCTCCCACTCTCTCCCGCGCGCA----- 4853
Qy 2135 euAlaGlnHisLeuSerGluValThrGlnAspTyrThrArgHisPro----- 2152
Db 4852 -----GACTCTTTCTTCATCACCCCGCAGGTG 4825
Qy 2153 -----G 2153
Db 4824 GACATGATCCCGCGAGGAGTAGTCGCTGGGAGACCGAGATCTCGGAGGCACTAC 4765
Qy 2153 lngInLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer---CysP 2172
Db 4764 CGCAGCGCTCGGACACACCTTAATCCCGCAGACCGCCACCTCATGCCAGATTGTG 4705
Qy 2172 ro----- 2172
Db 4704 CGCATGACTACGAGGACCTCATGCGAGGACACAACTATGATGTCACGTACGCAAC 4645
Qy 2173 --ValLeuAspLeuArgProProSerAspLeuTyrLeuProProAspHisGlyA 2192
Db 4644 GTCITTCGCGCGCGCGCGCGGAGCCATCGCATCGCTCGACGAGTGCATGGAG 4585
Qy 2192 laProAlaArgGlySerProHisSerGluGlyGlyLys-----ArgSerProG 2208
Db 4584 GACCTCGCGCG-----ACAAAGGGGGTCGCCAAGTCTTCACGCTTCCCC 4537
Qy 2208 luProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyLeuProValSerProp 2228
Db 4536 T-CCAGCTGCAG-----CCCGCTTCCCGCG 4511
Qy 2228 roGluGlyMetThr-----GluProGlyHisSerArgS 2239
Db 4510 CTCGACCGGCTACGCGCTCTCGTCGCTCCACAAATGAACCCCGCGGACCCAGAG 4451
Qy 2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlyS 2259
Db 4450 CGGCAACATCTCCA-----ACCTCAAGATCCAGG----- 4422
Qy 2259 erLysSerProGlyAsnThrSerGlnProProAlaPhePhe---SerLysLeuThrGluS 2278
Db 4421 -----CCAAAGTACCACATCATGTCCCGCTCATGACGCGCGCCGAGTCAACCGGTT 4370
Qy 2278 erAsnSerAlaMetValLysSerLysLysGlnGluLeuAsn-----LysLysLeuA 2295
Db 4369 CATC-AATGCCTACCAAGGCTCTGCCCAACCGCCATCAGCTGCTCTCCTCAGGACATTT 4311
Qy 2295 snThrHisAsnArgAsnGluProGluTyrAsnIle-Ser----- 2307
Db 4310 TCACCAACCGCGCAACGAGAGCGCTATGACTGGATCATCTACTGACCTGCCCGCGCG 4251
Qy 2308 ---GlnProGlyThrGluLeuPhe-----AsnMetProAlaIle--- 2319
Db 4250 ACCCAGCGCTCCAGTGGCTTTTACCTGCACCCCGCGAGGAGTGCATGCCCATGTATTCA 4191

Qy 2320 ThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGlnHisAlaSerThrAsn 2339
Db 4190 ACCTCCAAACACAGAGTCTACCGGTGCTCGAAGATACACAAACACCTGCAAGACAGAG 4131
Qy 2340 MetGlyLeuGluAlaIleAileArgLysAlaLeu-----MetGlyLysTyrAspGln--- 2356
Db 4130 AACGCTGCACAGCTATTATCGTTCCAAACATCCATAAGCTTTATTATTCTCAATAA 4071
Qy 2357 ---Trp-----GluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAla 2373
Db 4070 AGTTGGGCTGGAAAAGACAAACACCGCTCTCCATGGTCTTCTTCTCTCTCCCTCTCT 4011
Qy 2374 SerAlaSerLeuProAlaAlaMetPro----- 2382
Db 4010 TCCCCCAGTCCCGCTGCTAGTTGCTCCGCGCATGGACGCTCTCTCTGTCGCGCG 3951
Qy 2383 -----IleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSer 2396
Db 3950 TAGCTGAGCGCTTCGGGTAGTGTAGTAGCAGGACCGCAGGGTCTCTCTCCGGGTG 3891
Qy 2397 ProGlyGlyGlyLysAlaLysVal-----SerGlyArgPro----- 2409
Db 3890 GGTGGTGGTGGCGCTTCCACTCCAGAGGCTCAGGTCCTCCACACATCATC 3831
Qy 2410 -----SerSerArgLysAlaLysSerProAlaProGlyLeuAla 2422
Db 3830 GTCAGGACCATCTCATCTCTCTCTTCCATGGTGTTCGCTCCAGCGCGGCGCCACC 3771
Qy 2423 SerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArg 2442
Db 3770 TCTCTGTCACCAATCTCTCTCTCTCT----- 3741
Qy 2443 ThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer-ThrProPh 2462
Db 3740 -----TCTCTCTCGCGCTCCGAGGCGGTG 3714
Qy 2462 eProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProPr 2482
Db 3713 GCGC-----GCAGCTGTATCTCCACCCCTCC 3687
Qy 2482 oGlyLeuPro-----AlaGlySerGlyProLeuAlaGly-ProHisHisAla----- 2497
Db 3686 CGCAGCTCTCTGACGCTGCGCGAGCTCCGCCAGCGCTCCGCCATCATGCCACCTCC 3627
Qy 2498 -----TrpAspGluGluProLys 2503
Db 3626 GCGCTCAGATCCACCGATGGAACACCGCGCGAA 3592
RESULT 83
AAH23705
ID AAH23705 standard; DNA; 28906 BP.
XX AC AAH23705;
XX DT 08-AUG-2001 (first entry)
XX DE Human tumour suppressor gene 16, TSG16, genomic sequence.
XX KW Tumour suppressor gene 16; TSG16; human; immune response modulator;
XX KW inflammatory response modulator; signal transduction activator;
XX KW cytokine inhibitor; gene therapy; anticancer; anti-inflammatory;
XX KW autoimmune disorder; infection; chromosome 16q24.3; alternative splicing;
XX KW cellular proliferation suppressor; ds.
XX OS Homo sapiens.
XX FN WO200132861-A1.
XX PD 10-MAY-2001.
XX PF 30-OCT-2000; 2000WO-AU001329.

QY 1508 -----LeuLysSerArgProGlyThrAlaSer----- 1516
DB 7124 GGGAGCGCTTACGACATGCTCCAGGCGACCCCTGCTACTGCGCCCGAGCAGCTGCTCGTC 7065
QY 1517 -----SerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyL 1535
DB 7064 ACCGCCGCTCTGTCGCAAGCTGCAAGAGTCTACCGCGCTTCTGTCACACGAGGCC 7005
QY 1535 ysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuP 1555
DB 7004 GGCCTTC-CCACGCGCGCTTCAATGCTTCCAGAGGCCACCATCAGCT- 6956
QY 1555 roArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuS 1575
DB 6955 -----CCAACTCCACGCGCATCTTCAAGCAGCTCCT 6925
QY 1575 erSerLysAlaSerGlnAspArgLysLeuThrSerThrPro----- 1589
DB 6924 CTACCGCGCAGCGCGCCCGAGCGCGACCTCGGCGCGACCTCTCGGCGCCCTCCAC 6865
QY 1590 -----ArgGluIleAlaLysSerProHisSerT 1599
DB 6864 GAGCTGTATGACTACGTGCGAGAGAGCTGCGCGCGCGCTGCTACCCCAACTTCAT 6805
QY 1599 hr-----ValProG 1602
DB 6804 CGGCCCTCTCGGAGCGCCCTTACGCTATGACATCTGCGGCATGTATGCCAGCGCCT 6745
QY 1602 luHisPro-----HisProIleSerProTyrGluHisLeuLeuArgGlyV 1618
DB 6744 CACC-CACCCCTTCCCGCGCGCAACCCCTCAACCCCTTCGACCGCGCCCTGCGCGCC 6686
QY 1618 alSerGlyValAspLeuTyrArgSerHisIleProLeuAla-----PheAspProThrSerI 1637
DB 6685 GCCGCTGCGCAGGACCGCTCGACGCGCGCGAGCCCTCGACTACTTCGACCCCGACCTCC 6626
QY 1637 leProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaP 1657
DB 6625 TGCCCGCCCTCTTCCAGCATCGACGCGAC-----CCGC 6593
QY 1657 roAsnProThrTyrProHisLeuTyrProProTyrLeu----- 1669
DB 6592 CCAGCAGGACGAGCTGGAGCTCTCGCCCGCCCTACTGCTCCCGCAGGCGCGCGCTCT 6533
QY 1670 -----IleArgGly-----TyrProAspThrAlaAlaLeuG 1680
DB 6532 GCTGGACCAACGAGCCCTCCGCGCGAGTAGCCACCTCCGCGAGCTGCTCACCTCC 6473
QY 1680 luAsnArg----- 1682
DB 6472 ACAACCGTGGCTGGCGCTGCTGCGGACGAGCGCAACAGCATCTTCCCGAGT 6413
QY 1683 --GlnThrIleIleAsnAspTyrIle-ThrSerGlnGlnMethHisIleAsnThrAlaThr 1701
DB 6412 GGGCTGCTTCCCGCAGACTAGTCAGCTCAACATCGCTGCCAAGAAAGCGCGACC 6353
QY 1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlnSerSerLeuAla 1721
DB 6352 -----GGGACAGAAACCAACGCTCCGCGAGCATCG 6323
QY 1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741
DB 6322 CCAAGCTTCTCTCAAGCGCTCTACGGCAGTTTCG-----CCACAGCGCTCG 6275
QY 1742 ValLeu-----ValProThrProGlyThr 1750
DB 6274 ACARTAGAAAGTGGTCTTCGCGAGCCAGATGAGCGCGCGCACCCGAGCCATCGCCC 6215
QY 1751 ProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSer 1770
DB 6214 GCGCGCAGCTCT-----CCGTCCGCTCTCCGCTTCATAGAA 6176
QY 1771 ArgHisSerSer-----SerProLeuSerProGlyGlyProThrHisLeuThrLys 1787

DB 6175 CAGACAACCTCAGCGCTGAGGTGACCCCGGACTTTGTCTAGCGTACCCACCTGAAGACT 6116
QY 1788 ProThrThrThrSerSerGluArgGluArgAspArgAspArgGluArgAspArgAsp 1807
DB 6115 CCGGGACACACATGACGCGCGAGCGAGAGCGCGGAGACCGGAGACTCCGGTGACGAGG 6056
QY 1808 ArgGluArgGlu-----Lys 1812
DB 6055 CTTGCGAAGAACTGACGCGCCACGACCCCGCGCGCTTTTATACCCCGGACCGCG 5996
QY 1813 SerIleLeuThrSerThrThrThrValGluHisAlaProIleTrp----- 1827
DB 5995 AAGCCGCTCAGCTGGACTACCACTGGCGACCAATCCACTTCTCTGGACGCGGACGAGG 5936
QY 1828 -----ArgPro-----GlyThrGlu 1832
DB 5935 ACCTGTCTCCACAGCTCGAGAGCGCTCGCCATTAGTGCCCAACGACCGCTACCCCT 5876
QY 1833 GlnSerSerGlySerSerGlySerGly-----GlyGlyGlySer 1847
DB 5875 CGCAGGTGCGCAGCTTCTGCTCGCTGGCGCGCTTCTCTCCAGTGGCGCTCA 5816
QY 1848 SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThr 1867
DB 5815 TCTCTACCGCGACGACCGCGCATCCCATCGCGACCGCGCTCGCTCTCGCTACG 5756
QY 1868 -----GlnAspAlaLeuGlnArgProSerVal 1877
DB 5755 GGGACACCGACAGCTTCTCTCACCGCGCGGAGCGAGCTCATGGAGACCAAGGTA 5696
QY 1878 LeuHisAsnThrGlyMetLysGlyIle-----IleThrAla 1889
DB 5695 -----AACACCGCTCAAGCGAAGCGGGGCGCCCTCGTCTTCGATCCCAAGACCC 5642
QY 1890 ValGluPro-----SerLysProThrValLeuArgSerThrSerSerSer 1905
DB 5641 AGCTACCTCCTGGTGTGTCGAGTGGAGACCCAGTGGCGCGCTGCGCGAGCGCCACA 5582
QY 1906 ProValArgProAlaAlaThrPheProAlaThrHisCysPro----- 1920
DB 5581 GCTCCGAGAGCGTCTGCTCGCCCCCAGCTCTACGCCCTGCGCGACCGTTTGGATG 5522
QY 1921 LeuGlyGlyThrLeu-----AspGlyValTyrProThrLeuMetGluProValLeuLeu 1938
DB 5521 CTTGCGGACCGCGCGCCCGCAAGCTCCGCGCAAGGCGCATGCCACCAACACTCT 5462
QY 1939 Pro-----LysGluAlaProArgVal 1945
DB 5461 CCTATGACCTCTCCGCCAGTGTACTGAAACGACGCGAGCTCGGAAGCGCCACTTCA 5402
QY 1946 Ala-ArgProGluArgProAlaAspThrGlyHisAlaPheLeuAlaLysProAl 1965
DB 5401 GCACGAGCGCTCAGCTCAAGAGACCTGGCCACCGAGC-AGCCCCAGACGCGCCCC 5343
QY 1965 aArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuVa 1985
DB 5342 TTCACCGTCACGAGACCA-----CCCTCAGCGGG-----CCCTCCGG 5304
QY 1985 lProProValSerGlyHisAlaThrIleAlaargThrProAlaLysAsnLeuAlaProHi 2005
DB 5303 CCTTGAAGAGCGGACCTCCACC---CGCTCGACGCCCGCCCGCTCGTTCGCTACTCA 5247
QY 2005 shisAlaSerProAspProAlaProAlaSerAlaSerAspProHisArgGluLys 2025
DB 5246 C-----GCAGCAACCCCAACCGGAGAAC 5223
QY 2025 sThrGlnSerLysProPheSerIleGlnGluLeuArgSerLeuGlyTyrHis-- 2044
DB 5222 AACGACCTGCTGGAGGACGATACCGTGGGACATGTAACCGAGCTCTGGAGCGCATGC 5163
QY 2045 -----GlySerSer-TyrSerProGluGlyValGluProValS 2057

| | | | | | |
|----|------|--|---|--------|------|
| Db | 8577 | CCC----- | -----TGCCCCAGCTGCCCGCT-- | GAGACC | 8551 |
| Qy | 840 | GlyGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu | 859 | | |
| Db | 8550 | GGCCGCGCGCGCGAGG-CGG----- | | 8531 | |
| Qy | 860 | GluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAsp | 879 | | |
| Db | 8530 | ----- | CGGAGCGCGCACCGCGG----- | 8513 | |
| Qy | 880 | AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGly | 899 | | |
| Db | 8512 | GGCTCTGCTGAGCGCGCGCACCTCGCTCCAGGGCACACCCAGGACGGGAGA | 8453 | | |
| Qy | 900 | SerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAla | 919 | | |
| Db | 8452 | GCTGGAGCATCACTACACCGCCACTGGGACAGCGCCCTGGCAACCTCTTCAGCTGC | 8393 | | |
| Qy | 920 | ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer | 939 | | |
| Db | 8392 | ACTCTCGCGCGCGCGCGCCGACCTCTGCAGG----- | CTACCC | 8354 | |
| Qy | 940 | ProArgPro----- | SerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPro | 957 | |
| Db | 8353 | CCGACGCGGTGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 8294 | | |
| Qy | 958 | GlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAla-AlaIleProProlleGl | 977 | | |
| Db | 8293 | ----- | GCCAGCAACACCACTGCACC | 8273 | |
| Qy | 977 | nValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProPr | 997 | | |
| Db | 8272 | A---CGTGGCTGGGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 8228 | | |
| Qy | 997 | oAlaProProProGlnAenLynProGlnProGluSerAspAlaProGlnProGlySe | 1017 | | |
| Db | 8227 | CCTTCCCGCTCGCTGCT---CTTCACCCA----- | | 8200 | |
| Qy | 1017 | rSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAl | 1037 | | |
| Db | 8199 | ----- | GGGCGCGTGCACCT----- | 8185 | |
| Qy | 1037 | aGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVa | 1057 | | |
| Db | 8184 | -----CATCAGAGCATCCAGGAGGCCA----- | | 8161 | |
| Qy | 1057 | lProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTy | 1077 | | |
| Db | 8160 | -----GAAGTGGGAGCACTGGCGCGCATCTTCCAGACCGCCACACCTGCTCCG | 8111 | | |
| Qy | 1077 | rAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProVal----- | 1095 | | |
| Db | 8110 | TGCGCGCGCGGACTTCTACTTCCACCACTGTCAGCGCCAGCGCGCGCGCGCGCGCG | 8051 | | |
| Qy | 1096 | -----LeuProArgProProThrIleSerAsnProProProLe | 1108 | | |
| Db | 8050 | CCATCGCTTCGCGCGCATCGCTCCCGCA-----GGGCAACCGCGCGCTCT | 8003 | | |
| Qy | 1108 | uIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGl | 1128 | | |
| Db | 8002 | T----- | | 8002 | |
| Qy | 1128 | yMetSerValGlnLeuHisVal---ProTyrSerGluHisAlaLysAlaProValGlyPr | 1147 | | |
| Db | 8001 | -----CCTCCTATGACTTGGAGACCTACCTGGCAGCGCGCGCTTCGGAA | 7955 | | |
| Qy | 1147 | oValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa | 1167 | | |
| Db | 7954 | AGCAGCTGCTGCTCTCTGCTGCTCTCCACT----- | | 7921 | |
| Qy | 1167 | lLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValPr | 1187 | | |
| Db | 7920 | -----CTCGGGGACGAGGAGCTCTGGA----- | | 7897 | |

```
QY 263 Thr-ArgGlnTyrHisGluAsnIleLysIleAenGlnAlaMet----- 276
DB 10431 ACTGCGCCGCTTACAGGGCAATCCATTACACCGTGGAGCTCTTCCGCGCCGCTCGCA 10372
QY 277 -----ArgLysLysLeuLeuLeuTyrPheLysAr 286
DB 10371 ACATTGGAAACCGCGCCACCACTGACCGCGCCCTCCGCTC-----CG 10327
QY 286 gArgAsnHis-AlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGlnLeuM 306
DB 10326 CCGTCGCGCATCGC-----TGGATGTCGCGCTTCTACCTACCGCTACCAACCGCTCA 10276
QY 306 etGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAlaLysG 326
DB 10275 TG-----CTCATGAACCTCTCCCGCCGCCAAACCGGCCACGG 10240
QY 326 luSer----- 327
DB 10239 AACACTGGCCCTTACTCTACCTTCCGCGCCACCTGCTGCTGGGCTACCAACCTGC 10180
QY 328 -----LysValArgGluTyrTyrGluLysGlnPheProG 339
DB 10179 TGGCGCTCTGCAACGATTACATCTTTGACACGCGCGCTACAGTCAGCTGCGTACACCG 10120
QY 339 luIleArgLysGlnArgGluLeuGlnGluArgMetGln-----SerArgValG 355
DB 10119 AGGTC-----ATCCAGCCCACTACCAAGTGGTCAACTGGTCTGCTGCTGG 10075
QY 355 lGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluI 375
DB 10074 GCAACTGCTCTACACCATCAACAGCGCGCTTACCACCGGTTCTTGGACCTGGATGACT 10015
QY 375 leIleAspGlyLeuSerGluGlnGluAsn-----LeuGluLysGlnMetArgGlnL 392
DB 10014 TCTCAGACTCCCTGACGCGCATGACAGCGCCATCTTGGCGGAGCGGCTGGCGGACC 9955
QY 392 euAlaValIleProProMetLeu-----TyrAspAlaAsp----- 403
DB 9954 TGGCGTGTGTCGCCGCCCTCGCGGGTACGTCGCGCACCCACCTGGAGCGCGGAGCG 9895
QY 404 -----GlnGlnArgIleLysPheIleAsnMetAsnGlyLeu---- 415
DB 9894 TGCCCGTGGCACAATGCTGTCAGGAGCAGCTCAAGGACATCGCGCGCTGCGCAGCAGAGG 9835
QY 416 -----MetAlaAspProMetLysValTyr-----LysAspArgGlnValM 429
DB 9834 CTTGGGGCATGGCCGACCGCATCCGCTGACGCGCGCGCGCCCAAGACCTCACCTCC 9775
QY 429 etAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHis---- 447
DB 9774 TGGCGCCCATCCGCGAGCTCAAGACGGCTTACTTCTCTCTGTCAGCGCGCACTCC 9715
QY 448 -----ProLysAsnPheGlyLeu-----IleAlaSerPheL 458
DB 9714 GGTCCCGAGCACACCGCGCTCACTTACCTTGGCAGCTTGTCTGGTAGACGCTTCG 9655
QY 458 euGluArg----- 460
DB 9654 TCGAACGCTTCGCGGGTCCGAGAGCAGGTGCGCGGATTGTCGGAAGCGCTGCAACAGC 9595
QY 461 -----LysThrValAlaGluCysValLeuTyr----- 469
DB 9594 AGGAGGTACCTACGAGAGATTGACACGCTGCATCATCATGTCAGTGGCTCAGCCTCCCTCT 9535
QY 470 -----TyrTrleuThrLysLysAsnGluAsnTyrLys 481
DB 9534 GCAGCGCACCGCTGCGGGGGCGCTTTCGAGCTGCGGCCCGGAGCGCGCGCG 9475
QY 481 erLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGln-----GlnGlnG 499
DB 9474 CCGTACCGAGACATGCGCGCATGCGGGCGAGGTCTATCGAGCGCTTCGTGACCGGC 9415
QY 499 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
```


Db 5875 CGCAGGTGCGCAGCTTCGCTCGCTCGCTGGCGGCTGCTTCCTCGAGTGGCGCCTCA 5816
QY 1848 SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProArgThr 1867
Db 5815 TCCTCTACGCCGACGACCGCGGCGCCCAATCCCGACCGCCCTCGCTCGCTACG 5756
QY 1868 -----GlnAspAlaLeuGlnArgProSerVal 1877
Db 5755 GGGACACCGACAGCCTCTTCTCACCGCGCGGACCGAGCTCATGGAGACCAAGGTA 5696
QY 1878 LeuHisAsnThrGlyMetLysGlylle-----lleThrAla 1889
Db 5695 -----AACACCGCTCAAGCGGAAGCGGGGCCCTCTCGTCTTCGATCCCGACACCC 5642
QY 1890 ValGluPro-----SerLysProThrValLeuArgSerThrSerSer 1905
Db 5641 AGCTACCTGGCTGGTGGAGTGGAGACCAAGTGGAGCGCTCGCGAGGACGCCACA 5582
QY 1906 ProValArgProAlaAlaThrPheProAlaThrHisCysPro----- 1920
Db 5581 GCTCCGAGACGGTCTGCTCGCCCCCAAGCTCTACGCCCTGCGCGACACCGTTTGGATG 5522
QY 1921 LeuGlyGlyThrLeu-----AspGlyValTyProThrLeuMetGluProValLeu 1938
Db 5521 CTTGCGGACCGCGCGCGCGCAAGCTCGCGCCCAAGGCGCATGCCACCCACACCTCT 5462
QY 1939 Pro-----LysGluAlaProArgVal 1945
Db 5461 CCTATGACCTCTCCCGCAGTGTACTGACGACGCGGAGCTCGGAAGCGCCACCTTCA 5402
QY 1946 Ala-ArgProGluArgProAlaAlaAspThrGlyHisAlaPheLeuAlaLysProAl 1965
Db 5401 GCACAGCGCCTCAGCTCAAGAGACCGCTGGCCACCGAGC-AGCCCCAGACCGCCCC 5343
QY 1965 aArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuVa 1985
Db 5342 TTCACCGTCCACGAGACCA-----CCCTCAGCGG-----CCCTCCGG 5304
QY 1985 lProProValSerGlyHisAlaThrileAlaArgThrProAlaLysAsnLeuAlaProHi 2005
Db 5303 CCTGGAGGACCGGACCTCCAC-----CGCTCAGCGCCACCGCTCGTTCGTA 5247
QY 2005 sHisAlaSerProAspProProAlaProAlaProAlaSerAspProHisArgGlu 2025
Db 5246 C-----GCAGCAACCCCAACCGGAGAAC 5223
QY 2025 sThrGlnSerLysProPheSerileGlnGluLeuGluLeuArgSerLeuGlyTyHis-- 2044
Db 5222 AACGACAGCTGCTGGAGGACGATACCGTGGGACATGTAAACCGAGCTCTGGGAGCGCATGC 5163
QY 2045 -----GlySerSer-TyrSerProGluGlyValGluProValS 2057
Db 5162 GCGTGCTCCACGAGCGCTCGACGACATGCCCTACGCC-----GAGGCGCTCAAGCCCTCA 5106
QY 2057 erProValSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGlu 2077
Db 5105 AGCCTTGGCTCCCTCTCC-----CACCTCTCTCCC 5073
QY 2077 euAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys----- 2095
Db 5072 TGGGAGCGCGACACCTCTCGCAGGACCT-CGTCCAGAGAACCGCAACCTGGCGGACACC 5014
QY 2096 -----LeuGlyG 2098
Db 5013 CTCACGCTGCATGCCCTCTCTCACCGCGGTGGACCTGGCGCTCCCTCAACTACAG 4954
QY 2098 lyGluAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerP 2118
Db 4953 GTCCAGCCAGTCATCGCTGTCATCTACGGTCCCGACCGGCTCCCGGAGTCCCGAGCTGCTC 4894
QY 2118 ro-----LeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrL 2135

Db 4893 CGCAACCTCATCTCTCCCAACTCTCTCCCGCCCGCCGGA----- 4853
QY 2135 euAlaGlnHisIleSerGluValIleThrGlnAspTyThrArgHisHisPro----- 2152
Db 4852 -----GACTGTTTTTCTTCATCACCGCGCAGGTG 4825
QY 2153 -----G 2153
Db 4824 GACATGATCCCGCGCAGGAGGTAGTGGCTGGGAGACCCAGATCTGCGAGGCAACTAC 4765
QY 2153 lGlnLeuSerAlaProLeuProAlaProLeuTySerPheProGlyAlaSer--CysP 2172
Db 4764 CGCAGCGGTCCGGACCAACACCTTAATCCCCCAGACCGCCACCTCATGCCAGACTTTGTC 4705
QY 2172 ro----- 2172
Db 4704 CGCATGAGCTACCAGGACCTCATGCGAGGAGCACAACTATGATGTCACTGACTCACGCAAC 4645
QY 2173 --ValLeuAspLeuArgArgProProSerAspLeuTyLeuProProProAspHisGlyA 2192
Db 4644 GTCTTCGCGCGCGCGCGCGCGCGGACCATCGCCATCGTCTCGCAGAGTGCATGGAG 4585
QY 2192 laProAlaArgGlySerProHisSerGluGlyGlyLys-----ArgSerProG 2308
Db 4584 GACCTCGCGG-----ACAAAGGGGGTGCACAAGTTCTTCACCGCTTCCCC 4537
QY 2208 luProAsnLysThrSerValLeuGlyGlyGluAspGlylleGluProValSerProP 2228
Db 4536 T-CCAAGTGCAG-----CCCGCTTCCCCG 4511
QY 2228 roGluGlyMetThr-----GluProGlyHisSerArgS 2239
Db 4510 CTGCACCGGCTACCGCGTCTCGTCTCCACAAACATGAACCCCGCGCAGCAGAG 4451
QY 2239 erAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlyS 2259
Db 4450 CGCAACATCTCA-----ACCTCAAGATCCAG----- 4422
QY 2259 erLysSerProGlyAsnThrSerGlnProAlaPhePhe---SerLysLeuThrGluS 2278
Db 4421 -----CCAAGTACCACATCATGTCTCCCGCTCGAGCGCGCCAGTCAACCGTT 4370
QY 2278 erAsnSerAlaMetValLysSerLysLysGlnGluLeuAsn-----LysLysLeuA 2295
Db 4369 CATC-AATGCTACACCAAGGGTCTGCCACCGCCATCAGCTGCTCTCAAGGACATTT 4311
QY 2295 snThrHisAsnArgGlnProGluTyAsnIle-Ser----- 2307
Db 4310 TCACCCACCGCAACCAACCGCTATGACTGGATCATCTACTGCACCTGCCCGCCG 4251
QY 2308 ---GlnProGlyThrGluIlePhe-----AsnMetProAlaIle-- 2319
Db 4250 ACCGAGCCTCCAGTGGCTTTACCTGCACCCCGCGGAGGTATCATGCCATGTTTCA 4191
QY 2320 ThrGlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsn 2339
Db 4190 AGCTCAACACCGAGTCTACCGTGTCTCGAAGATACACAGACCCCTGCAAGACAGAG 4131
QY 2340 MetGlyLeuGluAlaIleIleArgLysAlaLeu-----MetGlyLysTyAspGln-- 2356
Db 4130 AAGCTGGAGACGCTATATCTGTTCCAAACATCATTAAGCTTTATTATTCTCAATAA 4071
QY 2357 ---Trp-----GluGluSerProLeuSerAlaAsnAlaPheAsnProLeuAsnAla 2373
Db 4070 AGTTGGCTGGAAAGACAAACACCGCTCTCCATGGTCTTCTTCTCTCTCTCTCTCT 4011
QY 2374 SerAlaSerLeuProAlaAlaMetPro----- 2382
Db 4010 TCCCCCAGACGTCGCGTCACTGTTCTCGCGCATGAGCCGCTCCTTGTCTCGCGCG 3951
QY 2383 -----lleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSer 2396
Db 3950 TAGCTGAGCGGCTTCGGGTAGTGGTAGCAGGACCGCGAGGTGCTCTCCGGGGTG 3891

| | | | |
|----|------|--|------|
| QY | 1267 | sValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValTh | 1287 |
| Db | 7659 | -----CCCCCTGGCCTA----- | 7648 |
| QY | 1287 | rGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaPr | 1307 |
| Db | 7647 | -----CCAGCTGGCGCGCGCGCTGGCGGCGCGCC | 7610 |
| QY | 1307 | oLysArgThrTyrAspMetMetGluClyArgValGlyArgAlaIleSerSerAlaSerIl | 1327 |
| Db | 7609 | -----CTTCCTCGAGCTTACATCGTGGCCACACATCTCGGGCT | 7568 |
| QY | 1327 | eGlu-----GlyLeuMetGlyArgAlaIleProProGluArgHisSer----- | 1341 |
| Db | 7567 | TCGATGAGATCGTCCTGGCGCGCAGGT--GCTCGCCACCGCGGAGGTGCCCCCT | 7511 |
| QY | 1342 | -----ProHis---HisLeuLysGluHisHisIleArgGlySerIleT | 1356 |
| Db | 7510 | CCTTCCTGGTCACCGCGCACCTTCATCGCGCAACGCGCGCATCTCTCAATGATCA | 7451 |
| QY | 1356 | hGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaL | 1376 |
| Db | 7450 | GCTTCAACCTGCCCAACCCCGCTTCGCGCGCGCGGACTACGGGACTGGAGGAGG | 7391 |
| QY | 1376 | ysLeuLeuLysArgGluGlyThrProProProProSerArgAspLeuThrGluA | 1396 |
| Db | 7390 | G-----CGTCATGGCCCCGAGGACCCCGCAGCACCATGATCGT----- | 7354 |
| QY | 1396 | laTyrLysThrGlnAlaLeuGlyProLeuLysIleLysProAlaHisGluGlyLeuValA | 1416 |
| Db | 7353 | -----CAAGTTATGTTGTCGGGACACCTTCGCCCT----- | 7324 |
| QY | 1416 | laThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisT | 1436 |
| Db | 7323 | -----CAGC----- | 7320 |
| QY | 1436 | hrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr- | 1455 |
| Db | 7319 | CACACACGCTCCGCAAGCGCGCGCGCTACGAGCTGGCCGT--GGACAAGGACAC | 7263 |
| QY | 1456 | --ProLeuLys-----TyrAspThrGlyAlaSerThrThrGlySerLysL | 1470 |
| Db | 7262 | TGCCCTTACGATGCGGTGAACGAGTTCTACCGCAGGGGGACTACTACGATCAGGAC | 7203 |
| QY | 1470 | ysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProL | 1490 |
| Db | 7202 | GGCTTCCACACGCGCAAGTACTGGGACAGCGATCAGGAGTACCTCTCTCAACAGGACCTG | 7143 |
| QY | 1490 | euAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer- | 1507 |
| Db | 7142 | TG-----GAAGAAGGAGAGCCAG | 7125 |
| QY | 1508 | -----LeuLysSerArgProGlyThrAlaSer----- | 1516 |
| Db | 7124 | GGGAGCGTTACGACATCGTCCAGGCCACCTCGCGTACTCGGCCACGAGGACGTCGTCGTC | 7065 |
| QY | 1517 | -----SerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyL | 1535 |
| Db | 7064 | ACCGCCGCTGTGGTCCGCAAGCTGCAAGAGTCTTACCGCGCGCTTCGTCCACACGAGGCC | 7005 |
| QY | 1535 | ysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuP | 1555 |
| Db | 7004 | GGCCTTC-CCACGCGCGCTTCAATGTCTTCAGAGGCCACCATCAGCT----- | 6956 |
| QY | 1555 | roArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeus | 1575 |
| Db | 6955 | -----CCAACCTCCACGCCATCTTCAAGCAGCTCT | 6925 |
| QY | 1575 | erSerSerLysAlaSerGlnAspArgLysLeuThrSerThrPro----- | 1589 |
| Db | 6924 | CTACCGCGCACCGCGGCCCGCGCGCATCTCGCGCAACCTCTCGCGCCCTCCAC | 6865 |

| | | | | |
|------|----|--|--|------|
| 1590 | QY | ---- | -----ArgGluIleAlaLysSerProHisSerT | 1599 |
| 6864 | DB | GAGCTGTATGACTPACGTCTCGCGAAGACCTGCGCGGCGCCCTCTCTACCCCAACTTCAT | 6805 | |
| 1599 | QY | hr----- | -----ValProG | 1602 |
| 6804 | DB | CGGCCTCTCGGGAGCCCTCTACGTCTATGACATCTCGGCATGTATGCCAGCGCCT | 6745 | |
| 1602 | QY | luHisPro----- | HisProIleSerProTyrGluHisIleuLeuArgGlyV | 1618 |
| 6744 | DB | CACC-CACCCCTTCCCGCGCGCAACCCCTCAACCCCTTCAGACCGCGCCCTTGGCGCGCC | 6686 | |
| 1618 | QY | alSerGlyValAspLeuTyrArgSerHisIleProLeuAla-- | PheAspProThrSerI | 1637 |
| 6685 | DB | GCOCCTGGCAGGACCGCCTCGACGCCCGCAGGCCCTCAGCTACTTCGACCCCGACCTCC | 6626 | |
| 1637 | QY | leProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisIleuAlaP | 1657 | |
| 6625 | DB | TGCCCGCCCTTTCACCATCGACGCCGAC | 6599 | |
| 1657 | QY | roAsnProThrTyrProHisLeuTyrProProTyrLeu----- | 1669 | |
| 6592 | DB | CCGACGAGGACCATGGACGTCTGCCCCCTTACTGCTCCCGCAAGGCGCGCGCTCT | 6533 | |
| 1670 | QY | -----IleArgGly----- | TyrProAspThrAlaAlaLeuG | 1680 |
| 6532 | DB | GCTGGACCAACGAGCCCTCGCGCGGAGTAGCACCTCCGTGGAGCGTGGTCAACCTCC | 6473 | |
| 1680 | QY | luAsnArg----- | ----- | 1682 |
| 6472 | DB | ACAACCGTGGTGGCGCTCGGCTCTGCGGACGAGCGCAACACATCTTCCCGAGT | 6413 | |
| 1683 | QY | --GlnThrIleAsnAspTyrIle--ThrSerGlnGlnMethHisAsnThrAlaThr | 1701 | |
| 6412 | DB | GCGCTGCTCGCCCGAGACTACTGTCAGCTCAACATCGTCCCAAGGAAAGGCGGACC | 6353 | |
| 1702 | QY | AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla | 1721 | |
| 6352 | DB | -----GGACAGAAACCAACAGCTCCGCGACATCG | 6323 | |
| 1722 | QY | LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro | 1741 | |
| 6322 | DB | CCAACTCTCTCCAAAGCCCTCTACGGCAGTTTCG----- | 6275 | |
| 1742 | QY | ValLeu----- | ValProProThrProGlyThr | 1750 |
| 6274 | DB | ACAATAAGAGGTGTCTTCCGCCAGATGGACGCCGACCAAGCCATCGCCC | 6215 | |
| 1751 | QY | ProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSer | 1770 | |
| 6214 | DB | GCGGCGACTCT----- | 6176 | |
| 1771 | QY | ArgHisSerSer----- | SerProLeuSerProGlyProThrHisLeuThrLys | 1787 |
| 6175 | DB | CAGACAACTCAGCGCTGAGTCAACCGGACTTTGTCGTAGCTACCCACCTGAAGACT | 6116 | |
| 1788 | QY | ProThrThrThrSerSerGluArgGluArgAspArgAspArgGluArgAspArgAsp | 1807 | |
| 6115 | DB | CCGGACCAACAGATGCAGCGCAGAGAGCGCGGAGCGCGGACTCCGCTGACGAGG | 6056 | |
| 1808 | QY | ArgGluArgGlu----- | -----Lys | 1812 |
| 6055 | DB | CCTGGCAGCAACTGACGGCCACGACCCCGCGCGCTTTTATACCCCGACCCGC | 5996 | |
| 1813 | QY | SerIleLeuThrSerThrThrValGluHisAlaProIleTrrp----- | 1827 | |
| 5995 | DB | AAGCGGTCACTGGACTACCACTGGCACAACCACTCTCTCGACCGGACGAGG | 5936 | |
| 1828 | QY | -----ArgPro----- | GlyThrGlu | 1832 |
| 5935 | DB | ACCTGTCTCCACACGCTCGAGAGGCTCGCCATTAGTGCCCAACGACCGCTACCCCT | 5876 | |
| 1833 | QY | GlnSerSerGlySerSerGlySerGly----- | GlyGlyGlyGlySer | 1847 |

Db 9273 CGCCCTCCTTAGAG----- 9260
QY 619 luMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaA 639
Db 9259 -----GAGGAGGTGGCGGCGCGCGCGAG-----GCCATCGCG 9223
QY 639 rgMetValGlySerLysThrValSerGlnCysLysAsnPhetyrPheAsn----- 655
Db 9222 CCCTGGAAGAGGAGCTCAGGTCCTCGCGCGCGCAGCAGCGAGTTTTCGCTTCGCCCGCG 9163
QY 656 -----TyrLysLysArgGlnAsnLeuAspGluIleLeuGlnHisLysLeuLysMetG 674
Db 9162 AGTTTACCGGGTCATGCAGCGCTCGAGGACCTGGCGGCGCATCAACGAGCTGACCGCTGC 9103
QY 674 lyls-----GluArgAsnAlaArgAGLysLysLysLysA 686
Db 9102 CGCGCTGGGTATAACTTCTCTCGCGGAGCAGATCGCCAC-CACCTCTCACTACCTG 9044
QY 686 laProAlaAlaLaserGluGluAlaAlaPheProProValValGluAspGluMetG 706
Db 9043 CACCAGCCTCGCCTGCAGGCGCCTTCAACCGCTGTGGCGCTCAACCTGGCCAG 8984
QY 706 luAlaSerGlyValSerGlyAsnGlu-----GluGluMetValGluGluA 721
Db 8983 GTAGTCATGCGCGCGGAGCGCGCAGCGCGAGCTGCTTACAGTCGCGTGTGGAACGAG 8924
QY 721 laGluAlaLeuHisAlaSerGlyAsnGluValPro-----ArgGlyG 735
Db 8923 AACGGAGAGAACCCCTGTGTGGCCATCATGAACCGCATCACCTCGACCTGGCGCCACC 8864
QY 735 luCysSerGlyProAlaIaThr-----ValAsnAsnSerSerAspThrGluSerIleP 752
Db 8863 GTGG-AGCGCGCGCGCACGGGACCTGGAGAGGAGCGAGCTCGAGAGTTTATGCAGGA 8805
QY 752 roSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProAlaIaT 772
Db 8804 CATCGCTACCGGAAACAGCGGGAGCGTGTCCGAGATCTCGCGCAGGTGGCCACCCA 8745
QY 772 hrLeu-----GlyAlaAspGlyProPro----- 780
Db 8744 CGATCGCGCATGATTCCGTGGAACTCTCTTCAGGTTCCGAGTCACCGGCGCGCTGGT 8685
QY 781 --GlyProProThrProArgThrSerArgAlaProIleGluProThrPro-Ala 799
Db 8684 CTTCTCGGGAACCCGAGATCCAGAACATCAACCGCGGTCTATCGGCCACGCTCGCG 8625
QY 800 SerGluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProPro 819
Db 8624 TCT-----GCGCGCGGACCACTCCCGCTGC-CCGAGCTCAACGCGCGCGTGC 8578
QY 820 ProValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGlu 839
Db 8577 CCC-----TGGCCAGCTGCGCGCT---GAGACC 8551
QY 840 GlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu 859
Db 8550 GGCGCGCGCGCGAGG-CGG----- 8531
QY 860 GluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAsp 879
Db 8530 -----CGAGGCGCGCAGCGCG----- 8513
QY 880 AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGlyGly 899
Db 8512 GGCTCTGTCTGAGCGCGCGCCACCTCGCGCTCCAGGGCACCACCGAGCGGAGGA 8453
QY 900 SerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAla 919
Db 8452 GCTGAGCATCACTTACCACCGCACTGGGACAGCGCCCTGGCCAACTCTTCCAGCTGC 8393
QY 920 ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer 939
|||

Db 8392 ACCTCTCGCGCGCGCGCGCGCGCTCTCTGCAGG-----CTACCC 8354
QY 940 ProArgPro-----SerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPro 957
Db 8353 CCCAGACCGCGGTGGAGCTGGCGCGCGCGCGCGCGCTCTCTCGTCCAGGTGGTCAACC 8294
QY 958 GlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAla-AlaIleProProIleG 977
Db 8293 -----GCCAGCAACACCACTGCACC 8273
QY 977 nValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro 997
Db 8272 A---CGTGGCTGGGACCCCGCGCGCGCGCGCGCGCGCG-----CCTCC 8228
QY 997 oAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnProGlySe 1017
Db 8227 CTTCCCTCTCGCTGCT---CTTCACCA----- 8200
QY 1017 rSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAla 1037
Db 8199 -----GGCGCGCTGCCACCT----- 8185
QY 1037 aGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVa 1057
Db 8184 ---CATCAAGAGCATCTCCAGAGGCCCA----- 8161
QY 1057 lProProArgGluValLysAlaSerProHisAlaProAspProSerAlaPheSerTy 1077
Db 8160 -----GAAGTCGAGCACTGGCGCGCATCTTCAGACCGCGCCACCTGCTCCG 8111
QY 1077 rAlaProProGlyHisProLeuGlyLeuHisAspThrAlaArgProVal----- 1095
Db 8110 TGGCGCGCGGAGCTTCTACTTCCACCGTTCAGCGCCAGCGCGCGCTGGTGGCAGA 8051
QY 1096 -----LeuProArgProProThrIleSerAsnProProIle 1108
Db 8050 CCATCGCTTCGCCCGCATCGCTCCCGCA-----GGCACCGCGCGCTCT 8003
QY 1108 uIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGln 1128
Db 8002 T----- 8002
QY 1128 yMetSerValGlnLeuHisVal---ProTyrSerGluHisAlaLysAlaProValGlyPr 1147
Db 8001 -----CCTCACCTATGCTGGAGACCTACACCTGGCAGCGCGCTTCGGAA 7955
QY 1147 oValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa 1167
Db 7954 AGCAGCTCGCTCTCTCTGCTGCTTCCACCT----- 7921
QY 1167 lLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValPr 1187
Db 7920 -----CTCGGGGAGAGAGCTGCTGGA----- 7897
QY 1187 oThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyLys 1207
Db 7896 -----CGCGCGCTGCAGGTGGTGGCGCTCCAGAGCGCG 7862
QY 1207 rIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgG 1227
Db 7861 CTGACGCTGGAGACCGCGCCAC----- 7834
QY 1227 ySerIleThrHisGlyThrProAlaAspValLeuTyrLys-----GlyThrI 1243
Db 7833 -----CTTCTACATGCTCTCGCGAGAGCGCGCCATCGCGCGCTTCAAGCGCTGC 7778
QY 1243 eThrArgIleIleGly-----GluAs 1250
Db 7777 GGCAGCGCTTCAGCGCGCGCTCACCGACACCTCTGGCAGCACCTGCTGGCCCAAC 7718
QY 1250 pSerProSerArgLeuAspArgGly-----ArgGluAspSerLeuProLysGlyHi 1267
Db 7717 CAGACCTCGCGAGCTGGCGGAGCGCGCGCGCGCTGCGGAGCGCGCGGGA-- 7660
|||

Db 4075 ATTCGGCGGTTCGTCATCATGACCAACGAATCGGTCCGGGTTCCTCGACTGCGCG 4016
QY 2255 Ser-----ArgMetGlySer 2259
Db 4015 AACGCAACAAACCCAGCAGCATGACACAGGTCCGGAACACACCTCGCGGGACCTG 3956
QY 2260 LysSerProGlyAsnThrSerGlnProAlaPhePheSerLysLeuThrGlu----- 2277
Db 3955 TCAGCCGCGCACCACACAGTTCACACACACAGGAGATCGCGGAATCGCGCGCGCCCA 3896
QY 2278 -----SerAsnSerAlaMetValLysSerLys-----LysGlnGluLe 2290
Db 3895 ACCAATCCTGGATCAGTCAAGATCGTGTGTGTCGCGCAGCTGCTCGGAAGCAACGT 3836
QY 2291 AsnLysLysLeuAsnThrHisAsnArgAsnGlnProGlyAsnLeuSerGlnProGly 2310
Db 3835 CCAAGTCTCGCGAGACGACCCATCGCATGCGACACACAG-----CGTCCGGG 3788
QY 2311 ThrGluLeuPheAsnMetProAlaLeuThrGlyLeuMetThrTyrArgSerGln 2330
Db 3787 GCG-----GGTCAGCAACCCGTCAGTTCGGGTGG-AGTTCGGTCCCAAGTGCC 3738
QY 2331 AlaValGlnGluHisAsnThrAsnMet-----GlyLeuGlu 2343
Db 3737 GAGCCGAGGCTCGCGCGAGTTCGTATTCGCGTCACTGCCAGGAGACCGCTTGGA 3678
QY 2344 Ala-----llelleArg 2347
Db 3677 GCGCGCGGAGATTCGTTCCAGTCGACCCGGAACAGGACTCGTGTTCATCGCAGG 3618
QY 2348 LysAlaLeuMetGlyLysTyrAspGlnTrpGluLys----- 2360
Db 3617 GCGCTGCCAATCTGTCGAAATCTGCGCAGCAGCAGCAATCATGAGAGCAGG 3558
QY 2361 -----ProLeuSerAla-----AsnAlaPheAsnProLeuAsnAlaSer 2374
Db 3557 GCTTGGCCAGTCTGTCGTCGTGAACACCGCAGCATTCGCGCGCGGCAACGAGG 3498
QY 2375 AlaSerLeuProAlaAlaMetProLeuThrAlaAlaAspGlyArgSerAspHisThrLeu 2394
Db 3497 TGACCCGGAAGGAAGTGGCGCGTCTGTTGAGGACACACCTTGCACGCGCAACGGG 3438
QY 2395 ThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
Db 3437 AGCATCCCTCGTCGATCTCT-----TCGCTCGCGCCAGCGCTCAGTGCC 3387
QY 2415 LysSer-----ProAlaProGlyLeuAlaSerGlyAsp 2425
Db 3386 GCATCCAGCAGCAGGCTGGATTCGGAACCTCGCGCGTCTTCGCGGTGTCGGTGGG 3327
QY 2426 ArgProProSer-----ValSerValHisSerGluGlyAspCysAsnArgArgThr 2443
Db 3326 AGCCCGACCTCGCGGAAGATCTCGTTCGCGGCTTCACGCGCTCGCAAC----- 3276
QY 2444 ProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhePro 2463
Db 3275 -----CCCTGGAACGCGCGGCCGAATCGAAGCCCGCTCAGCTGCGCG 3231
QY 2464 TyrAsnProLeuMetLeuGlnAlaGlyValMetAlaSerProProProGly 2483
Db 3230 GCCTA-GAAGCTTCAGCGCAACCGCAGCATCGTTGGGCGGCAATCCCATCGG 3172
QY 2484 LeuProAlaGly-----SerGlyProLeuAlaGlyPro 2494
Db 3171 GCGCTGCGCGGCTCTCGTGTGATCGCTGAGGTCAACGTCGCGTCCGCGGAGGTCCA 3112

RESULT 81

AAZ30163/c

ID AAZ30163 standard; DNA; 34094 BP.

XX AAZ30163;

AC AAZ30163;

XX

DT 17-OCT-2003 (revised)
DT 26-JAN-2000 (first entry)
XX Complete nucleotide sequence of the PAV-3 genome.
DE
XX PAV-3; defective recombinant PAV vector; live recombinant virus;
KW subunit vaccine; nucleic acid immunisation; gene therapy;
KW genetic disease; hemophilia; cystic fibrosis; cancer; viral infection;
KW acquired immune deficiency syndrome; PAV antigen; porcine pathogen; da.
XX Porcine adenovirus; Type 3.
OS
XX WO953047-A2.
PN
XX 21-OCT-1999.
PD
XX 15-APR-1999; 99WO-US008220.
PF
XX 15-APR-1998; 98US-0081882P.
PR (UYSA-) UNIV SASKATCHEWAN.
XX Reddy PS, Tikoo SK, Babiuk LA;
PI WPI; 1999-620422/53.
DR
XX New nucleic acids from the genome of porcine adenovirus-3, and derived
PT gene therapy vectors, particularly for immunization.
PT
PS Example 2; Fig 1; 87pp; English.
XX The present sequence represents the complete nucleotide sequence of the
CC genome of porcine adenovirus-3 (PAV-3). The specification also describes
CC a defective recombinant PAV vector comprising inverted terminal repeats
CC (ITR), packaging sequences and at least one heterologous nucleotide
CC sequence (II), but lacking EI functions. The defective vectors replicate
CC inefficiently in cells (other than helper cells) so are unlikely to be
CC immunogenic. Deletion of the EI (and optionally other regions) increases
CC the size of heterologous insert that can be packaged. The PAV-3
CC polynucleotide sequences are used to produce (recombinant or defective)
CC vectors that can express heterologous proteins, e.g. for making live,
CC recombinant virus or subunit vaccines, for nucleic acid immunisation or
CC for gene therapy (e.g. of genetic diseases such as hemophilia or cystic
CC fibrosis, cancer, or viral infections, including acquired immune
CC deficiency syndrome), also for in vitro expression of recombinant
CC antigens (for antibody production), antisense RNA, ribozymes or
CC therapeutic proteins. They are also used diagnostically to detect PAV
CC antigens and/or nucleic acid. The vectors may be used in human or
CC veterinary medicine, but particularly for expressing protective
CC determinants of porcine pathogens. Regulatory regions may be used to
CC control expression of heterologous genes. Antibodies raised against PAV-3
CC polypeptides can also be used for diagnosis (to detect PAV-specific
CC antigen). (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 34094 BP; 6240 A; 11070 C; 10693 G; 6091 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1-03e-09 Length: 34094
Score: 606.00 Matches: 655
Percent Similarity: 30.98% Conservative: 305
Best Local Similarity: 21.14% Mismatches: 992
Query Match: 4.59% Indels: 1159
DB: 2 Gaps: 143
US-09-522-753-5 (1-2517) x AAZ30163 (1-34094)
QY 22 HisSerLeuSerTyrProValGlnLeuAlaArgThrHisThrAspValGlyLeuLeuGlu 41
Db 11264 CACAGGTCCACGAGCAGCGGTCTCGGGCGCGCCAGGCTCAGCGCTCTT--- 11208
QY 42 TyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIle----- 58
Db 11207 -----CACCATGCC-----CTCGTCCCGCAGTGTCTGCGCATCAG 11172

| | | | | | | | |
|----|------|--|------|----|------|--|------|
| QY | 1667 | ProTyrLeuIleArgGlyTyr---ProAspThrAlaAlaLeuGluAsnArgGlnThrIle | 1685 | QY | 1998 | ProAlaLysAsnLeuAlaProHisHisAla---SerProAspPro | 2011 |
| Db | 5939 | CCGCAAGAGCCAGATACACTCACCAGACCGAGTGTGACCCGCCAGGTGCAGCGCC | 5880 | Db | 5101 | CCAGA-----CGCCGAACATGCTCGCGCAACCAACCAACAGCTGGTTCCT | 5051 |
| QY | 1686 | IleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGln | 1705 | QY | 2012 | ---ProAlaProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro | 2030 |
| Db | 5879 | ACCAACGACGA-CGAACA---CGCGTATCCACCGTCACCGCAGGACCCCTCGAAACCAA | 5824 | Db | 5050 | GATCCAGTCCGGCAGCTCCGACGCGA-CCAACACAGACCGTTGACGTCCGCGACTCG | 4992 |
| QY | 1706 | ArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAla | 1725 | QY | 2031 | PheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerTyrSerPro | 2050 |
| Db | 5823 | CGA-----ATACGCAACCCGACCCGAAACACGCTCCCGCCTGC-CATACCTA | 5774 | Db | 4991 | TTGACC-----GCGCTCCGCGCACCGGCGCTGGCACCAGGTCCGCAACAA | 4944 |
| QY | 1726 | AlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValPro | 1745 | QY | 2051 | GluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeu | 2070 |
| Db | 5773 | GATNAACCTCGAAC-----CCT | 5756 | Db | 4943 | CGAGCACACCGCTCGCGCGC-ATCAGCGCGCAGGCGGTTAGGTGACGTGACGACGG | 4885 |
| QY | 1746 | ProThrPro----- | 1748 | QY | 2071 | ---ProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys | 2089 |
| Db | 5755 | CCGCGCCCTCGTGATGAACCGCGCGCGTAGTGTGTGTCATCAGCCCGCAACACGC | 5696 | Db | 4884 | AACCAAGACGATTCCCGCGAACCGCAACCGCATCTCGAA | 4843 |
| QY | 1749 | ---GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln | 1766 | QY | 2090 | GlnProGlyProValLysLeuGly----- | 2097 |
| Db | 5695 | CGTCCGCTCCCCCGCAACGAAGCGGATCAATCCCGCCGCTCCACCGTTTCCCAAG | 5636 | Db | 4842 | CAGCCGCAACCGCTCCGCGCGACATCGCGGATTGTCGCGCAATCGCGCTACGCC | 4783 |
| QY | 1767 | ProPheSerSerArgHisSerSerProLeuSerPro-----GlyGlyProThr | 1783 | QY | 2098 | -----GlyGluAlaHisLeuProHisLeuArgProLeuProLysSer | 2112 |
| Db | 5635 | CGACTTCAGCAGCAACCGCTGCTCGGATCCATCGCAACCGCTCAACCGGGGAAATCC | 5576 | Db | 4782 | CGCGCATCGACGCGCTCCGCTCATCCGCTGGGTGCTCCCAACAGCCCAAGCAGCGA | 4723 |
| QY | 1784 | -----HisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAsp | 1799 | QY | 2113 | -----GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHis | 2129 |
| Db | 5575 | CGAAGAAACCGGCATCGAACTCTCGGCATCTCGAGGAGCCACCTCACACATACG | 5516 | Db | 4722 | CACCGCAGCAGCCCGCGCTCCGCGGTGTGAGCCAGCGCTCCAGGAACGCTTCGC | 4663 |
| QY | 1800 | ArgAspArgGluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSerThr | 1819 | QY | 2130 | GlnArgValValThrLeu----- | 2135 |
| Db | 5515 | AGTCCCAAGAC---GATCGGATCCGATCGA-----ACA | 5483 | Db | 4662 | CGCGGTAGTTTCCTGTACCCCGCAGCCCGCGAGAACGAGAGAGAGACAC | 4603 |
| QY | 1820 | ThrValGluHisAlaProIleThrArgProGlyThrGluGlnSerSerGlySerGly | 1839 | QY | 2135 | ----- | 2135 |
| Db | 5482 | ACCCCGCTGATCCCGCAGCAGCATCTCGCGAACTCCGCGCATTCGAAGCCCGCG | 5423 | Db | 4602 | GAAGAACGAGAGATCCAGATCAAGATCAACTGTGTGAGTGCAGTCCAGCCCGGCCTT | 4543 |
| QY | 1840 | SerSerGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln | 1859 | QY | 2136 | ---AlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGln | 2154 |
| Db | 5422 | CCACCA-----AACGCCAAGCTCTCGCGGAGGACACCCACCGGATAAC | 5375 | Db | 4542 | CGGCGCGAACACCTTTGCCACGTGTCTGTG-TGTGACGCGACTCGACGACCCCGTGGCAA | 4484 |
| QY | 1860 | HisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHis | 1879 | QY | 2155 | LeuSerAlaProLeuProAlaProLeu-----TyrSer | 2165 |
| Db | 5374 | GACAGCCCA---TCCCCACGATCAACACGGGATCTGACAGTCAACCG- | 5330 | Db | 4483 | GTGCCCCGCGCG---CATGCACCACTCCGCGCAGCGGAACTTGTCCGAGATGCTGCA | 4427 |
| QY | 1880 | AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg | 1899 | QY | 2166 | PheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeu | 2185 |
| Db | 5329 | ---CGCCACACCGCGGTGCGGAGTCCGAGCTCCACGCCCAATCATGTTGTGCA | 5273 | Db | 4426 | GCAGCGGTCCAACTGGGCGCGATCCCGACATCGCAGTCCGACCTCGACCGCAGCTC | 4367 |
| QY | 1900 | SerThrSerThrSerProValArgProAlaAlaThrPhePro-----ProAlaThr | 1917 | QY | 2186 | Pro---ProProAspHis-----GlyAlaProAlaArgGlySer | 2197 |
| Db | 5272 | GCAAGAACCGAACCAACCGCGGTGTCGATAGTCGAACACCGCAGTAGCGCGCAACC | 5213 | Db | 4366 | CCAGGCCCACTCATCCACAGTGTGTGACCCCGGGGCGCTCCAGCCGCTCGCC | 4307 |
| QY | 1918 | HisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu | 1937 | QY | 2198 | ProHisSerGlu---GlyGlyLysArgSerProGluProAsn----- | 2210 |
| Db | 5212 | GAACGCCCA----- | 5204 | Db | 4306 | CCGCGACACCGCGCGGACACCGCTCCCGCAACATCCCGGGCAACCAACCCGC | 4247 |
| QY | 1938 | LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis | 1957 | QY | 2211 | -----LysThrSer-----ValLeuGlyGly | 2217 |
| Db | 5203 | -----GCACCCCGCCAGCCGTTCCGAACCTCCACCGCGAACAGATCAA | 5156 | Db | 4246 | CCAGACCGCGTACCCCGCAATCAACACCGCTCCATCCGACCGGCACTGGGATCC | 4187 |
| QY | 1958 | AlaPheLeuAlaLysProAlaArgSerGlyLeuGluProAlaSerSerProSerLys | 1977 | QY | 2218 | GlyGluAspGlyIleGluProValSerProGluGlyMetThrGluProGlyHisSer | 2237 |
| Db | 5155 | -----AACCCAAATCCCGAAAG | 5138 | Db | 4186 | GGCTGTCTCTCGCGCGCAACACATCGCGCAGCGGAGCA-----CCCGACTTCAC | 4136 |
| QY | 1978 | GlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThr | 1997 | QY | 2238 | ArgSerAlaValTyrProLeuLeuTyrArgAsp----- | 2248 |
| Db | 5137 | CACGATCGGCCCGACCTCGACCGACCCG-----AATGCC | 5102 | Db | 4135 | CGGCGCGCAACCGCAGCTCGGTTCCTCGGAACCGCAGCGCGCGCGCAGCGCATG | 4076 |
| | | | | Db | 2249 | -----GlyGluGlnThrGluPro | 2254 |

Db 7905 ACAGCGTGTGCGCGCGCCACCGCGGACACAGTTTCGGTTTCGTACGGTCTTTACG 7846
QY 1090 spThrAlaArgProValLeuLeuProArgProThrThrIleSerAanProPro----- 1106
Db 7845 CAGCATCGGAATCGGATCGCTGATCAGCCCAAGATGCAATCGGAATCAGGGCCGA 7786
QY 1107 -----ProLeuIleSerSerAla----- 1112
Db 7785 CAGCGTCGCATCAGGCGGAGTTTGAAGATCGTGGCGACCCCTTGTGCGGAGCACCG 7726
QY 1112 ----- 1112
Db 7725 GACGCGTCGCGAAGCGGACGGCTCTCGAACCCTGACGCCACCAATACTCTCGCGGTAGC 7666
QY 1113 -----LyHisProSerValLeuGluArgGlnIleGlyAlaIle 1126
Db 7665 CATCACACCTCGCGTGTGAGCTCACCGTCAACGTGACACACCGGTAGGT----- 7613
QY 1126 erGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLaPro----- 1144
Db 7612 -----CGCCACGCGGTAGTCCACGCT 7591
QY 1145 -----ValGlyProValThrMetGlyLeuProLeuProMetAspProLysL 1160
Db 7590 GCGTGGGATTTCCGCGAACTCTCTGCAACATCGGT-----CCATCCGATACGAATGAACGC 7534
QY 1160 ysLeuAlaProPheSerGlyValLysGlnGlnLeuSerPro-----ArgGlyG 1177
Db 7533 ATCGGAACCGCAACCGCGCCACCAAGATCCCTTGCCTCCATCCAGCGCGCGCGATGTC 7474
QY 1177 lnaAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgG 1197
Db 7473 GTCAAGCACCTCCGATCGCGGAGAGCACCCGACTCAGGAG-----CGTT 7426
QY 1197 lyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgV 1217
Db 7425 G-ACCSCG-----GGATACCGCGCCCATCA---ACCGATCGGCCAACAGCGCGCA 7376
QY 1217 alProSerAspSerAlaIleThrTyr-----ArgGlySerIle 1230
Db 7375 GCTGCTCTCCACAGCAGCCAGCCGCAACATGCGCAGCGCCAGGAAGCAAGCCCTGATCA 7316
QY 1230 hrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleGlyGluA 1250
Db 7315 ACCGGCAGCAGCCGCG-----ACCATCCGAGCGCATCCGGCAACG 7274
QY 1250 spSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIle 1270
Db 7273 ACAACAC-TCCCGCGCAACCGCGCGCGGAAT-CCGCGACCGAATGGCCCGCACCCAC 7216
QY 1270 yrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCys- 1289
Db 7215 ACCC-GCCCGGACACCCACGAACCCAGC-----AGGCTCAGCAACCAACCTGCA 7166
QY 1290 -----SerLysGluAspGly-----ArgSerSerGlyPro-----ProH 1302
Db 7165 ACGCGAACAACCCGACTGGGCGCCACAGCGTCCGATCCCAACAGGACCCGTCGGAACCAA 7106
QY 1302 isGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAla 1322
Db 7105 ACACCAATCTCCGAGACCCCACTCTGCTGCCCAAGATGCGCGTCCAGTCGCGCATGCTT 7046
QY 1322 leSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGlu----- 1338
Db 7045 CGTCAACAGCGTCAGC-GAACACCGGAACCAACCGAGTAGGCCCCCGGCCATCCCAAC 6987
QY 1339 -----ArgHisSerProHisLysLysGlu-----GlnH 1349
Db 6986 CACTGTCCGCGCTGACCGGGAACACGAACCCCAACCCCGCAGAGACCGGGATCCCGAC 6927
QY 1349 iHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGlu 1369
|||||

Db 6926 ACCAC-----CCCAAGACCTCGCGACCGGAGCCAGC 6894
QY 1369 spTyrLeuArgArgGluAlaLysLeuLysLeuArgGluGlyThrProProProProp 1389
Db 6893 GCTTCCAAACCCGACCAACCAAGTTTCGGCTCGCTCCCA-GCACCACCGCGCTCCTC 6835
QY 1389 roSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysP 1409
Db 6834 CAACGCGCGCGGAGGAACCAACGAAACG-----C 6802
QY 1409 ro-AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIle 1428
Db 6801 CGATCCACGCTGGAATATCGCCGCTACCCAAATAGGACAAATCGCGCTCGCCTG 6742
QY 1429 ProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlu 1448
Db 6741 AGCGCTGAGCGTGCCTGCTCTCCCGACAGCATCAACGGGACCGCTGGAAGTCCAC 6682
QY 1449 GlySerIleThrGlnGlyThrProLeuLysTyrAsp----- 1460
Db 6681 GGTGATTCGGAATCGGACCGCTCTGTCCGAGTCTGCTGTGTGTTTTCGAGGCTTG 6622
QY 1461 -----ThrGlyAlaSerThrThrGlySerLys---LysHisAspVal 1473
Db 6621 TTCGAGGATGACGTGTGCTGTGCTGCTGATCCGAAACGACGACACGCCCGCCGCG 6562
QY 1474 ArgSerLeuIleGlySerProGlyArgThrPhe-----ProPro 1486
Db 6561 AGGATGGCTGTTCGCGGCGCAGGGGTGTCT-CCGTGAGGAGCTGGACATCCCGCGCG 6503
QY 1487 ValHisProLeuAspVal-----MetAlaAspAlaArgAlaLeuGluArgAla 1502
Db 6502 ACCAATCACTTCCGACGTAGTTTCATCCATGCAACGTGCGGCGAGCTGCCGTGCC 6443
QY 1503 -----CysTyrGluGlu 1506
Db 6442 GCATGCCATACCATTTGTATCACACGAGCAGACCCCGCGCGCTCGTATGACCAA 6383
QY 1507 ---SerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGly 1525
Db 6382 TATTCAGCTTCCGACCCCAACCAACGCGCGATCGCTATCCCGCCCTGCCGT--- 6326
QY 1526 AlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAsp 1545
Db 6325 -----AAGTGCATACAGAGCTCGCGCTCAATCGAT 6293
QY 1546 HisGlyAlaProPheAlaGlyHisLeu-----ProArgGlySerProValThr 1561
Db 6292 CACCAAGCTGTGTTCCAGTCCATCGCGCTCCACGCGCTCCACATCAGACACCGCAAC 6233
QY 1562 MetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGln 1581
Db 6232 CAGCACTCCCAACGCT-----GGTGATCACCGCTGCTCGAGGGCCATTCG 6182
QY 1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db 6181 -----GCGCGCTCAACCATTCGACG----- 6161
QY 1602 GluHisHisPro-----HisProLysSerProTyrGluHisLeuLeuArgGly 1617
Db 6160 ---CACATCTGTATTCACCGCTACCCCGCGACGACCGCCGAGAACCGCGTCCGCTGC 6104
QY 1618 ValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIle 1637
Db 6103 GCCGG-----CATCCGACGCGCTCCCAACCAACACC-AAGCCGCGCTTCT 6057
QY 1638 ProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
Db 6056 CCCCACCGGTGCA---TCCGCGCGCGCAACAGACTTGACCGCCCATCCCGCGCC 6000
QY 1658 AsnPro-----ThrTyrPro-----HisLeuTyrPro 1666
5999 AAGCCCGCTTGACGCGAATACTCCAGAACATCCCGCGGTGGCCATCACCGTACACCA 5940


```
Qy 532 uGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLys-ThrAspA 552
Db 9952 TATCAGAGTCAGCTGTCGAGATTCCAGCCCGCCCGCTTCGGCGGAACACCGCTC 9893
Qy 552 spThrGlyGluAspAsnAspGluLysGluAlaValAlaSer 566
Db 9892 CCACGTCGTCGCTGACCAAGCTCGATCACCCCGTCGCGCAGCCCGCCGCGCATGCA 9833
Qy 567 -----LysGlyArgLysThrAlaAsnSerGlnGlyArgGlyGly 581
Db 9832 CCACGCGCCAAACGGAAATTCGCGCGGACCGCCCGCCAGCAACAGCTCCAGATCAGCAC 9773
Qy 581 rgileThrArgSerMetAlaAsnGluAlaAsnSer 593
Db 9772 GATCCGCGAGCTGCACCGCCACACCGTCGACCGCAGCGCGCGCTTCAGCTCACCCCA 9713
Qy 593 luGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer- 612
Db 9712 CCAACTCGGTGATTCCAGGAGCATCCAGCGCACGCGCCCGTCAACACCAATCGCGCGGA 9653
Qy 613 -----SerArgTrp-----ThrGluGluG 619
Db 9652 CACCGCTTCGCTCACCAAGTGGCGGCAACCGCGCCCGCCCAACAGCGCCGTCGCGCCG 9593
Qy 619 luMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg----- 632
Db 9592 AAATCAGCACGTTTCGGTC-GGG--ATCCGCCCAACACCGCGCGGACGTCGATCTGCCCG 9537
Qy 633 --AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnP 652
Db 9536 GCGACGAAGCAACAGCGGGCGCTGCGACCGCGCCCTTCGCGACGGCCAACTCGGG--T 9480
Qy 652 heTyrPheAsnTrpLysLysArgGlnAsnLeuAspGluIle-----LeuG 667
Db 9479 TCCATCGAGCCCAACGCACTCGGCAACCGCACCGCATTCGCGCGTTCGCTCGACGTC 9420
Qy 667 lnglnHisLysLeuLysMetGluLysGluArgAsn-AlaArgArgLysLysLysLysAla 686
Db 9419 GCCAGCACAACTGCCCCGGTTCCTCCACCTGCGCGGCGACCAACACCCCAACACGCG 9360
Qy 687 ProAlaAlaIleSerGluGluAlaAlaPheProValValGluAspGluGluMetGlu 706
Db 9359 GCATGTCGAGCTCTGATGACCGCTACTCGGAGCGCTGTC-----ACC 9315
Qy 707 AlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAla 726
Db 9314 GCTGCACAGTCACCAACCAACAGAGCTGCTCGGCCAACCGCGGCTCAGCCCAACAGCT 9255
Qy 727 -----SerGlyAsnGluValProArgGly 734
Db 9254 TGGAGCAGAGCCAAAGCTGTCGGGTCTCCGAAGTGGCGTGCCTCGCGGTGCGGGATCC 9195
Qy 735 GluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPro 754
Db 9194 GCCTGTTCT---CCGCGCAGTAGACCAACGACATCAGGAACAGCGCGCTGGAATCAAAT 9138
Qy 755 HisThrGluAlaAla-LysAspThr----- 762
Db 9137 TTCTCGGTAGCGCGCCAGATCCGCGCAGCGTCCGCGGAATCCGTTCCGAGTTTGATG 9078
Qy 763 -----GlyGlnAsnGlyProLysProProAlaThrLeuG 774
Db 9077 AGATCGTCTTTACCGGAGGAGTATCACCAGCAGGCGCCAGCGGCTTCGCTCGAC 9018
Qy 774 yAlaAspGlyProProGlyProProThrProThrProArgThrSerArg-----Al 792
Db 9017 TGGCTGCCACCGCTCCAGGTCACTCGAACAGCGCTCGCGGACGACGCTCCCGATTC 8958
Qy 792 aProIleGluProThrPro-----AlaSerGluAlaThrGlyAl 805
Db 8957 GCCAGTGTCTCCGACGACACGCTTCGTAGCTGAGAGATCGATTCGAGCGGACCAACCG 8898
```

```
805 aProThrProProAlaProPro-----SerProSerAlaPr 818
8897 CCATCTCATCATCAGCGCACCGATATCGCTCGGTGCGCTCGCTCTCCAGCGCAC 8838
818 oProProValProLysGluGluLysGluGluThrAlaAlaAlaProProValG 838
8837 CTG-CTGATCGTCTCGCGAGCGGAA-----AGTTCCGACGCGGTGAA 8794
838 uGluGlyGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAl 858
8793 TGAACCGCAGCCACCGCGCGCTTCTCTCC-----GG 8761
858 aGluGluProValLysSerGluCysThr---GluGluAlaGluGluGlyProAlaLysG 877
8760 GCGGTCCGCGAGCGCGGAGCTCGACGCGGCATCCAGCAACGCTGGTTCGACAGTA 8701
877 yLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysG 897
8700 TCCCGACGCC---ATCGCATCGCTCGCGCGCGGCGATTTCCGGCAGAACCTCGTC 8644
897 uGlyGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGlnAsp----- 914
8643 ACCAGCGCATCGCTGAGCCCTGAGCCCTGGACACACGCGCGTACTCTGAACCGGTTTC 8584
915 -----SerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
8583 GCGCAGCGAGCTAGAACTCATCGAGCAATCGACTCGGCACCTCAGCGCAACATC 8524
929 a-Glu-----GlyGlyAspL 934
8523 AGGAACCGCCCTACCAACCGGTGTCGCGACGAGTTCGCGGTGGGTGGCGGTCCA 8464
934 ysAsnArgLeuLeuSerProArgPro-----SerLeuLeuThrProThrGlyAsp 950
8463 ACCCCACCTCTGACCCACCGCGCGGCGGGAATACAGCGACACCTTCGCGCACCGCC 8404
951 -----ProArgAlaAsnAla 956
8403 TTCTCGGTGCGGTTCACCAACGAGCTGGATTTTCGATGGCGTCTGCTCGCCACAGTTC 8344
956 exProGlnLysProLeuAspLeuLysGln----- 968
8343 TTCGCGAGTTTCAATTTAAACTGTGTCTCTCCGCGAGGCGCACGACTACCGCACCGCC 8284
968 ----- 968
8283 CGCGCGTTCGGGACCAACCGCGTGTTCACAGGTGAGCTCTTCACCGGTTCACAGCC 8224
969 -----ArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluP 984
8223 GAGACGCGCTCGTGCAGGGCCATCTCCAGATGGCGGTCCCGGGACGACGACCTC 8164
984 roProArgGluAspAlaAlaPro---ThrLysProAlaProPro-----A 998
8163 GCCCAGCAGCGATGCTGGSCCAGCCAGCATCGTCCCACCAACACCGCGGTTCAG 8104
998 laProPro-----ProGlnAsnLeuGlnProGluSerAspAlaProGlnG 1014
8103 CACCATCCCATCACCATCGCGCACCAACCGCGCCCCCAGCAACGAGTCTCCGCGC 8044
1014 lnProGlySerSerProArgGlyLysSerArgSerProAlaProAlaAspLysGluA 1034
8043 AGCCAGCGCGATACCTCCATCGC---CGTAATCGATGTTTCCAGCCAGTACCGCTG 7987
1034 laPheAlaAlaGluAla-GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
7986 CCGCTGGAAGCATACGTCGCGCACTC-----GACCCGCGTGC----- 7948
1054 ProPhePro-ValProProArgGluValIleLysAlaSerProHisAla-ProAspPro 1073
7947 -----CCCGGTGCG-----GCGAATACGCGCACCATTCGACCGGACACC 7906
1073 erAlaPheSerTyrAlaProProGlyHisPro-----LeuProLeuGlyLeuHisA 1090
```


PS Claim 7a; Page 284-313; 354pp; German.

XX This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
 CC identify, inactivate or modulate genes involved in the biosynthesis of
 CC (iii); (ii) to generate a library of polypeptide synthases; (iii) for
 CC adding forosamine or trimethylthamose to a spinosyn or polyketide
 CC aglycone; and (iv) for recombinant production of the corresponding
 CC enzymes, which are used for production of (II), their precursors or
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (ii) are
 CC macrolides with insecticidal, but not antibacterial, activity, and can
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence, ORF 22, encodes an S. spinosa polyketide synthase

XX SQ Sequence 16767 BP; 2256 A; 4805 C; 6457 G; 3249 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5,02e-10 Length: 16767
 Score: 608.50 Matches: 680
 Percent Similarity: 32.10% Conservative: 317
 Best Local Similarity: 21.89% Mismatches: 1171
 Query Match: 4.60% Indels: 958
 DB: 4 Gaps: 139

US-09-522-753-5 (1-2517) x AAF88339 (1-16767)

QY 16 GluProArgTyrProPro-----HisSerLeuSerTyrProValGlnIleAla 31
 DB 11458 GAACCA-----CCGCCAACACCGGATGCCATTCGCGGAGCATCGACAAACCGCTCCA 11405
 QY 32 ArgThrHisThrAspValGlyLeuLeuGluTyrGlnHis---HisSerArgAspTyrAla 50
 DB 11404 ACAACACCAAC-----CAGCACCTCACCACCCCAACCGGTACCG- 11367
 QY 51 SerHisLeuSerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSer 70
 DB 11366 TCAGCGCTCTCCGGAACGACTTGACCGCCGCTCGCGTCCAGGCGCCCG-CTGACGGGA 11308
 QY 71 GluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSer 90
 DB 11307 GAATTC-----CACAAAGCTCCCGGCGTGCACAT 11278
 QY 91 TyrLeuProGluLeuGlyLysSerGluMetGluPhe----- 102
 DB 11277 CACCGTCACACCAACCGGCAAGGCGGAGATCGCATTCGCGGAACGCAAGCACTGACCCGC 11218
 QY 103 ----- 103
 DB 11217 CAATGCAACGCCCAACGACGACGACGACGAGTATCCCGTACCGGAGGACCCCTC 11158
 QY 104 GluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu 123
 DB 11157 GAACCAACAGTACGC-----AATCGGCGCGACGCGACACTCCGAGC 11113
 QY 124 LeuAla-----ThrGlyGlnProAlaGly 131
 DB 11112 ACTGCCATTGCCGAGATACCCCTCGAAACCTTCGCGGGCTCGGCTGCGGAACCGGCGCGC 11053
 QY 132 SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProPro 151
 DB 11052 ATAGTCGTGTGATCATGACCCCGCGCAACACACCGGTCG-----ACCTCCCT 11005
 QY 152 SerProHisThrAspPro---GluLeuGluLeuValProArgLeuSerLysGlu 170
 DB 11004 CACGGAACCGGATCGATCCCTCCCGCTCAATGGCTCCAGGACAGCTCCAGCAACAG 10945
 QY 171 GluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnIle 190
 DB 11012 GACGAGGAGCGCGGCAACACCGGAAAGCGAGGAGCAACACCGGAAGACGACAGTCCG 9953

DB 10944 CCGTTGCTG-----CGGATCCATCGC-----CAACGCTC 10915
 QY 191 SerLysLeuLysLysGlnGlnLeuGluGluAlaAlaLysProGluPro 210
 DB 10914 ACGGGGGAATCCCGAAGAACCCGATCGAACCGATCGCGTGGTGGAGGAGCCCC 10855
 QY 211 GluLysProValSer---ProProProlleGluSerLysHisArgSerLeuValGlnIle 239
 DB 10854 TTCCCGCAGTAGTCTTCCAGCCGCTTC----- 10825
 QY 230 IleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGly 249
 DB 10824 -----CGGTCCGATCGAACAGCTGCAGATCCAGCCGCGATCGGC 10780
 QY 250 -----ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265
 DB 10779 TGGGAAGTCGCGCAGCATCCAGCCCCCGGCCACCAACGCGCAGCTCTCCGCGGA 10720
 QY 266 TyrHisGluAsnIleLysIle-----AsnGlnAlaMetArgLysLysLeuLeuTyr 283
 DB 10719 GGACACCTGACCGGGAACCGCATCCCATCCGACTATCGGACGAGATCGTCTTCGC 10660
 QY 284 PheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAsp 303
 DB 10659 GACCGGTGCCGTGTCGTCCAGCGCGCTGCGT-----CGTAGCGAT 10615
 QY 304 GlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsnProArgArgArg 323
 DB 10614 CTCGTCTGCAGTCTCTG-----GTGAGGAACCGGCGCGCTCGGCTCGG 10564
 QY 324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhePro---GluIleArgLys 342
 DB 10563 ATAGTCGAACACCGCAGTCCGCGCAACCGCACCCCGCAGCACCCGCGGCGTTCG 10504
 QY 343 GlnArgGluLeuGlnCluArgMetGlnSerArgValGlyGlnArgLysSerGlyLeuSer 362
 DB 10503 CAACTCCACACGAGACACGAATCAAAACCCAGATCCCGGAAAGCGCGTTCGACCCCGAC 10444
 QY 363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382
 DB 10443 CTCAGCGCGCTCGAATACCCAGCAGCGCGCGCATGCTCCCGCACACCCCGAGCAG 10384
 QY 383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValleProProMetLeuTyrAspAla 402
 DB 10383 GGTGCCCTCGTCCGCGAGCTCGGTTCGCGTAGC-----CCGCTATTACGACGATGGG 10330
 QY 403 AspGlnGln-ArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422
 DB 10329 GGTACCAAGTCCGTC-----AACATGCGGCGGCGAGCGCGACCGCGC 10285
 QY 422 lTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPheArgG 442
 DB 10284 CTGCTCGGCGAAGTCTCTGTCNACTTCGCC-----GGAACACTCCGCTGC 10237
 QY 442 uLysPheMetGlnHisPro-----LysAsnPheG 452
 DB 10236 CCGGTCCAGCGCGCAGCGCGCATCGAAACGCGCAATCTCTCTGTTGGTGAACACGCGCG 10177
 QY 452 yLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrIle 472
 DB 10176 GTTGGCGCGCGATCTCTGCTCGGTCGACCTC-GCTGAG-----CGTCT 10133
 QY 472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGlyLys 492
 DB 10132 CGTCACTCCGCTCGGCTCTCCACAGCGCGCGCGAGCGACGCGAGCGAGTCCCT 10073
 QY 492 sSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 512
 DB 10072 GTGACCGCGGTAAATGCGCAATGCGTCAGGAACGCGTTCGCGCGCGCATTAATTCGCT 10013
 QY 512 tProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluLysGluLysG 532
 DB 10012 GACGAGGAGCGCGGCAACACCGGAAAGCGAGGAGCAACACCGGAAGACGACAGTCCG 9953

Db 20883 C-GAGCACCTCCCGCAGTCCCAACCCGCGGAGAGTCGCGGAGCAGCTGCGGCGACGCGGCC 20941
Qy 1482 rgThr-----PheProValHisProLeuAspValMetAlaAspA 1496
Db 20942 ATGCTGTGCGGAAGGCGGCGCTCCCGCCA--CAACCGCTCGGCCACGCGGTCAG 20998
Qy 1496 laArgAlaLeuGlu-----ArgAlaCysT 1504
Db 20999 TGGCGCCCTTGACCGGGAACACAGGACCACCGCCGCGTGACGTGCGCGCGCGCTC 21058
Qy 1504 yGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaA 1524
Db 21059 ACAGCGCGCG-----CCGCGCTCACACCCCGCAGGCGGT-CGAGCCCGCC 21105
Qy 1524 rgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrG 1544
Db 21106 GAGCAGCGCT-----CCGCGTCGCGGCGACACAGCGCGCGCTCGCGCTG 21156
Qy 1544 luAspHisGlyAlaProPheAlaGlyHisLeuProArgGly-----SerProV 1560
Db 21157 GCGCGGAGAGCAGCAGCGCGCGCGCGACGTCTCTCGGGCGCAGCGCGCGCTCCCG 21216
Qy 1560 alThrMetArgGluProThrPro----- 1567
Db 21217 CAGGTCCCGCCAGCGCGCGCGCTGGGGCGCAGACCGCGCGTGTGCGCGCCCGACAC 21276
Qy 1567 ----- 1567
Db 21277 GACGACCGGACGCGGCGCGGACGCGGACAGGACACGCGGCGCATCGGCGCGCAGCGG 21336
Qy 1568 -----ArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgL 1584
Db 21337 CAGCGCGCAGAGGCGCAGCGCGCTCGCGCGGAAGTTTCGCGGCGCGCGCTCACGCC 21396
Qy 1584 yLeuThrSerThr----- 1588
Db 21397 CGGTGCGCCACCGCGCGCGCGCGGAGGTGAGCGCGTGTGCGCGGCTGCCACGCGCTC 21456
Qy 1589 -----P 1589
Db 21457 CGCCACGCGCTCCGCCACGCGGCGCTCTTCGAGGATCAGTGTGGCGTGTGGTGGCTGGC 21516
Qy 1589 roArgGluLeuAlaLysSerProHisSerThrValProGluHisHisPro----- 1605
Db 21517 CCGAAGGCGCAGACCGCGCGCGCTCAGCCGACCGCGCGCGCGCGCGCGCTCC 21576
Qy 1606 -----HisProIleSerProTyrGluHis-LeuLeuArgGly 1617
Db 21577 ACCCGCACGCGCGCGGAGCACCCGCTGCGGCGCGCGCGCGCGCGCGCTCCAGCC 21636
Qy 1618 ValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIle 1637
Db 21637 ACCAGCAGCATCCCGAGTGCCTCCTCGCTCCCG-----TCCCGCAG----- 21681
Qy 1638 ProArg-----GlyIleProLeuAspAla----- 1645
Db 21682 CCAAGCAGGCGCGGTCGAGCACCGCGCACCGCGCGCGCTCCAGTGCACCTCGGTGGT 21741
Qy 1646 -----AlaAlaAlaTyrTyrLeuProArgHis 1654
Db 21742 CGCGTGTGCGGCGTGGAGGTGGGGGCGAGGACCGCTTGGCAGCGCGCTGCACCACT- 21800
Qy 1655 LeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrPro 1674
Db 21801 TCAGCACCGCGCGCACCGCGCGCG-----CCT 21830
Qy 1675 AspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694
Db 21831 GCGTGTGCGCGATGTTCCAGTCTCACCGACCCGAGCCACAGCGCGCGCGCGCGAC 21890
Qy 1695 MetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSer 1714
Db 21891 CGCGCGCGTAGGTCTTCAGCAGCGCGCTGCGCTCGATCG-----GGT 21932

Qy 1715 ProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleLeuAspLeu 1734
Db 21933 CCCCAGCAGCGTCCCGGTCCGCTCGCTCGCAGCAGGTCCAGT----- 21977
Qy 1735 SerGlnValProHisLeuProValLeuValProProThr----- 1747
Db 21978 -----CCGCGCGCAGACAC----- 22010
Qy 1748 ProGlyThrProAlaThrAlaMetAspArgLeuAla-----TyrLeu 1761
Db 22011 CCGGATCACCGCTGCTGCGGGGACCGCTGGGCGCGGTGAGCCCGTTGGACGCGCGT 22070
Qy 1762 ProThrAlaPro-GlnProPheSerSerArgHisSerSerSerProLeuSerProGlyG 1781
Db 22071 CGGAGTTACCGCGCGCGCGCACCGCACCGCAACACCGCGCGCGCTCCCGCGGCGG 22130
Qy 1781 yProThrHisLeuThrLysProThrThrThrSerSerSerGluArg----- 1796
Db 22131 CGGAGAGCGGTTTCAGCAGCAGCACCGCGCGCTCCCGCGCGGTGCGCTCCGCGT 22190
Qy 1797 -----GluArgAspArgAspArgGluArgAs 1805
Db 22191 CGCGGAGAACCGCTTGCAGCGCGCTCGCGGCGCAGCGCGCTCCCGGAGAACTCGA 22250
Qy 1805 pArgAspArgGluArg-GluLysSerIleLeuThrSerThrThrThrValGluHisAlaP 1825
Db 22251 CGAACGTGTCCGCGCTCGACATCACCGTCACTCCCGCGCGCGCGCATCGAGCACTCGC 22310
Qy 1825 rolleThrArgProGlyThr-----GluGlnSer----- 1834
Db 22311 CGCGTGGAGCGCGCGCACCGAGGTGATCCGACACGACGACGAGCGCGGTGT 22370
Qy 1835 -----SerGlySerSerGly----- 1839
Db 22371 CCACCGTATCGGCGCGCTTCAGCGCGAGGTGTAGCGCGCGCGCTCCGATCGCGCGC 22430
Qy 1840 -----SerSerGlyGlyGlyGlySerSerSera 1850
Db 22431 TGCAGGAGTGTGCGGTCCCGAGTAGCTCCAGCTCCGCGCGCAGCGCGCGAACCGCG 22490
Qy 1850 rgProAlaSerHisSerHisAlaHisGlnHisSerProIleSer---ProArgThrGlnA 1869
Db 22491 AGCGGT 22550
Qy 1869 sp---AlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIle 1888
Db 22551 CGGGGCGATCGCGCGCTCGAACCGCTCCACGCGACCTCCAGCAGCAGCGG-GTGC 22609
Qy 1888 hrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerProVala 1908
Db 22610 TCGGGT 22668
Qy 1908 rgProAlaAlaThrPheProAlaAlaThrHisCysProLeuGlyGlyThrLeuAspGlyV 1928
Db 22669 GGCCA-----TGTCGCGCAGGA----- 22685
Qy 1928 alTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgP 1948
Db 22686 -----ACGCGCGTCCCGCACCTAGGAGCGCGCGAC 22716
Qy 1948 rgGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
Db 22717 CCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 22758
Qy 1968 lLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProV 1988
Db 22759 C-CGCTGCTGGGAACCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22817
Qy 1988 alSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlas 2008
Db 22818 CTTGCGCGCAGGTGACC-----CCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 22868

| | | | |
|----|-------|--|-------|
| Qy | 915 | erAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysA | 935 |
| Db | 19075 | CA:::---CGTC | 19080 |
| Qy | 935 | sNArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnA | 955 |
| Db | 19081 | GCGGGGGTGGGGCTCCGGCAGCT-----CCACCGGGGCCACGCGGTCTCCAG | 19131 |
| Qy | 955 | laSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProp | 975 |
| Db | 19132 | CAGACC-----CGGTGGAC-----CGAGCCGGATCCCGCCCC | 19167 |
| Qy | 975 | rolleGlnValThrLysValHisGluProProArgGluAspAlaAlaPro | 991 |
| Db | 19168 | CC-----GTGGGAAGCCAGTCCCGCATCCGGCCACCGGGCGGGCGGC | 19212 |
| Qy | 992 | -----ThrLysProAlaProProAlaProProp | 1001 |
| Db | 19213 | GCGTCCGGCAGGTCCGAGCAGACGCCCGTCCCGACCGCGCGCGCCACCCGCAC | 19272 |
| Qy | 1001 | roProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg | 1021 |
| Db | 19273 | CG-----CGTCCACGCGGAGCGTGTCAGCGGACGCG | 19305 |
| Qy | 1021 | lyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnL | 1041 |
| Db | 19306 | GGAGAAGAGAACCGCACCCACCGCGCGCGT----- | 19340 |
| Qy | 1041 | ysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArgG | 1061 |
| Db | 19341 | -----CCCCCGGCAC-----CCCCCGGCAC | 19350 |
| Qy | 1061 | luValIleLys-----AlaSerProHisAlaProAspProSerAlaPheSertYrAla- | 1078 |
| Db | 19351 | CCGTGCGCGGTGAAGTGTCTACGTGACGCGGCACCTCCAGCAGCGCGGTGCAGCGC | 19410 |
| Qy | 1079 | ProProGlyHisPro---LeuProLeuGlyLeuHisAspThrAlaArgProValLeuPro | 1097 |
| Db | 19411 | GAACCGCGCGCCCTGCGCTGCTCGCGGTGACCTCGACCTCGCCGAACACCTCGTCGC | 19470 |
| Qy | 1098 | ArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerVal | 1117 |
| Db | 19471 | GCGCGCCACACCGCGGTGAC-----CCGCTGGA-AGTCCGCGCGTACTCCAGCGCCT | 19523 |
| Qy | 1118 | LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTy | 1137 |
| Db | 19524 | TGGCGCGGAATCCGGGTAGAGCCCGCGAGGTGACCGGTTCG----- | 19568 |
| Qy | 1138 | SerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp | 1157 |
| Db | 19569 | -----CGCCCGCGGGCGCCACTGGC-----CCCCGACCGGTCCG | 19604 |
| Qy | 1158 | ProLys-----LysLeuAlaProPheSerGlyValLysGln | 1169 |
| Db | 19605 | GCGCGCGGGCGGGCCGAGGACGCGCGCGGTGCAGACCG-----GGGGTGGTCCG | 19658 |
| Qy | 1170 | GluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAla | 1189 |
| Db | 19659 | CGGAGCAGCATCTCACGGGGCGGTCCGTCCGCGCGCGGGCCGACACCGACGCA | 19718 |
| Qy | 1190 | GlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThr | 1209 |
| Db | 19719 | CC-GCCAGTCAACCGTCTCCGGGACCAACACGCGCGCGCCACACCTCTCTCCAG | 19777 |
| Qy | 1210 | LysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyArgGlySerIle | 1229 |
| Db | 19778 | GCGGGCAGCCG--ACCTCGTCCCGCGCGCACCGCC-----ACCTCC | 19819 |
| Qy | 1230 | ThrHisGlyThrProAlaAspValLeuTyIlysGlyThrIle----- | 1245 |
| Db | 19820 | ACCAGCGCACACCG-----GGCAGCAGCGCGCGCGCACCTCTCTCCAG | 19861 |
| Qy | 1246 | IleIleGlyLysAspSerProSerArgLeuAspArgGlyArgGlu----- | 1260 |

[illegible]

| | | | |
|------------------------|--------|---------------|------|
| Percent Similarity: | 30.35% | Conservative: | 276 |
| Best Local Similarity: | 21.36% | Mismatches: | 1033 |
| Query Match: | 4.61% | Indels: | 1120 |
| DB: | 7 | Gaps: | 135 |

| | | | |
|---|-------|--|----------------------|
| US-09-522-753-5 (1-2517) x AAL61224 (1-82746) | | | |
| Qy | 13 | ArgAlaThrGluProArgTyrPro | -----ProHisSerLeu 24 |
| Db | 16506 | CGGGCCACGCGGCTGCTCGGCCACGACGCGGACACCGTCCACCTCGACCTCGGGGTGG | 16565 |
| Qy | 25 | SerTyrPro-ValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHi | 44 |
| Db | 16566 | GCCTGTCACGTGCAGGTGCGCGCAGCACCCTGCGGAGGCGCTGCACGGTCTTCA | 16625 |
| Qy | 44 | sHisSerArgAspTyrAlaSerHisLeuSerProGlySerIlelle | 59 |
| Db | 16626 | CCACACCGCGAT-----CCCGCGCGCGCTGCATGTGCCGAGGT | 16667 |
| Qy | 60 | -----GlnProGlnArg | 63 |
| Db | 16668 | TGGACTTGGAGGACCCAGCACGACGGCTCCCGTCCCGTCCCGCTAGGTGGCGA | 16727 |
| Qy | 64 | -ArgArgProSerLeu-----LeuSerGluPheGlnProGlyAsnGluArgSerGlnGl | 81 |
| Db | 16728 | CCAGCGCGGGCTCGATCGGTGCTGCCAACCGGCTCCCGTCCCGCTCCACCG | 16787 |
| Qy | 81 | uLeuHisLeuArgProGlnSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGl | 101 |
| Db | 16788 | CGTCCACGTCGA-----AGTGGACAGAC | 16811 |
| Qy | 101 | uPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSe | 121 |
| Db | 16812 | CGGCTCGCGAGCGCGCCG-----GATCACCGCTCTCGCGCGCAC | 16856 |
| Qy | 121 | rProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLe | 141 |
| Db | 16857 | C----- | 16857 |
| Qy | 141 | uThrGlyLysLeuGluProValSerProPro-----SerProProHisTh | 156 |
| Db | 16858 | -GCTCGGCGCGTCAACCGTTTCAGCCGCCGTCGAGTTCACCGGCTCCCGCACCA | 16916 |
| Qy | 156 | rAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuGlnAsnMe | 176 |
| Db | 16917 | CGGCCAGCACCGCGCGCTTGGCGCGCGCTC----- | 16950 |
| Qy | 176 | tAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLy | 196 |
| Db | 16950 | ----- | 16950 |
| Qy | 196 | sGlnGlnGlnLeuGluGluAlaAlaLysProProGluProGluLysProValSerPr | 216 |
| Db | 16951 | -CGACAGCGCTCCAGCAGCACCAACCGCACCTCGGACACCGCTCCGCTCCGCTCCG | 17007 |
| Qy | 216 | oProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLy | 236 |
| Db | 17008 | -----CGACGCCGAGA | 17011 |
| Qy | 236 | sLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluProLe | 256 |
| Db | 17019 | ACGGCTTCAGCGCGCGTCCG----- | 17044 |
| Qy | 256 | uTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMe | 276 |
| Db | 17041 | ----GGCGAGCGCGCTCGCTGGCGACTCCAGAACCC----- | 17077 |
| Qy | 276 | tArgLysLysLeuIleLeuTyrPheLysArgAsnHisAlaArgLysGlnTrpLysGl | 296 |
| Db | 17077 | -----CCTGGGCGTGGCCATGACCGTCCGCGCGCGCCCA | 17111 |
| Qy | 296 | nLysPheCysGlnArgTyrAspGlnLeuMetGluLeuGluLysLysValGluArg-I | 316 |

| | | | | | |
|-------|----|-------|-------------------------------|-------------------------------------|-------|
| 10870 | Db | GC | CGAGGCGCAGCTGTAGAGCCGGTACAA | CGGCTCTTTGAGCGCGCGGATGATCT | 10811 |
| 1541 | Qy | u | ---ThrTyrGluAspHisGlyAla | Pro-PheAlaGlyHisLeu-ProArgGlySerPro | 1559 |
| 10810 | Db | CC | ACAGATACCAGACTACTGTGACCCAGC | GTGTCGAGTCAGATGTCCTACCTGGCCGC | 10751 |
| 1560 | Qy | Val | ThrMetArgGluProThrProArgLeu | GlnGluGlySerLeuSerSerSerIysAla | 1579 |
| 10750 | Db | ACT | TCTGTCTTCGACACACCCACCGGCA | CGCAGCAAGACGAGGCAAGGCGAAATGAGC | 10691 |
| 1580 | Qy | Ser | GlnAspArgGlyLeuThrSerThrPro | ArgGluIleAlaLysSerProHisSerThr | 1599 |
| 10690 | Db | GCC | CAG-----GCCGACACGGCCAC | CCGCCGGAAGG | 10655 |
| 1600 | Qy | Val | ProGluHisHisProHisProIleSer | ProTyrGluHisLeuArgGlyValSer | 1619 |
| 10654 | Db | AC | CGCGGCCACGCTGCCAC----- | CACCTCGTGAAGTGTCCGC | 10613 |
| 1620 | Qy | Gly | Val-----AspLeuTyrArgSerHis | IlePLeuAlaPheAspProThrSer | 1636 |
| 10612 | Db | CGT | GTCTCTACCAAAACCGCGCTCCGCG | CTCGATTCCTCCGCGCACCGCG | 10553 |
| 1637 | Qy | Ile | ProArgGlyIleProLeuAspAlaAla | -AlaTyrTyrLeuPro----- | 1652 |
| 10552 | Db | GTG | TGTGCGGTGCTGTCGCGGTGTCG | TGGCGGTGGTTCGCCCAAGCGGC | 10493 |
| 1653 | Qy | ----- | ArgHisLeuAlaProAsnProThrTyr | ProHisLeuTyrProProTyrIleuI | 1670 |
| 10492 | Db | CTG | ACGAGCTGCGCTACACCGCGGCAC | GCATCACCATGTCACCGTCAAGCCCACTTAT | 10433 |
| 1670 | Qy | e | Arg-----GlyTyrPro-AspThrAla | alaLeuGluAsnArgGlnThrIleAla | 1687 |
| 10432 | Db | CAC | GGGTCAATTGGAATCCGCGACGCG | CGCACTGGATGAAGCGGCCACCACTCG | 10373 |
| 1687 | Qy | sn | AspTyr---IleThrSerGlnGluMet | HisHisAsnThrAlaThrAlaMetAlaGlnA | 1706 |
| 10372 | Db | ATG | ATCTTCGTGTTTCATCGCGCGCAG | CTACAC-----GCCGTTC | 10333 |
| 1706 | Qy | rg | AlaAspMetLeuArgGlyLeuSerPro | ArgGluSerSerSerLeuAlaLeuAsnTyrAlaA | 1726 |
| 10333 | Db | CGC | GTGCTGCTGCTTCCGCG----- | CCACGACGGCG----- | 10303 |
| 1726 | Qy | la | GlyProArgGlyIleIleAsp----- | ----- | 1733 |
| 10301 | Db | ----- | CGTGGTGTGTGATGTTTGGGCGGT | CGCATCCCGGAATCTTGTCTGAAG | 10255 |
| 1734 | Qy | ----- | LeuSerGlnValProHisLeuProVal | ----- | 1742 |
| 10249 | Db | ATG | TGTGGCGGCGCGCGCTCGGT | CGGGTTCGCTGTACCTGTTGCTGGGTGG | 1019 |
| 1743 | Qy | ----- | LeuValProPro----- | -----ThrProGlyThrProAlaThrAlaA | 1755 |
| 10189 | Db | GTG | GCGGTCTGGTACA-CCGCGAC | KATCTGTGCACACGCGGGGTGACCCGCTGGTGT | 1013 |
| 1755 | Qy | et | AspArgLeuAlaTyrLeuProThrAla | ProGlnProPheSerSerArgHisSerSerS | 1775 |
| 10130 | Db | GCT | GTGTTGCGCGCGGTGTACAGCAT | CGCGGCAATCTCTACGCGTACGCT---- | 1007 |
| 1775 | Qy | er | ProLeuSerProGlyIleProThr----- | HisLeuThrLysProThrThrThrSerS | 1793 |
| 10074 | Db | ----- | GGCCGACCGCTGGCGCAGCAGCT | TCGGCTACCAACGAGTTCCT | 1003 |
| 1793 | Qy | er | SerGluArgGluArgAspArg-Asp | ArgGluArgAspArg----- | 1807 |
| 10031 | Db | CC | ACGCTGCACCGCGTGGCGGATCTG | CCACTACATCCGATGTTGTTGTTGTTGTT | 9972 |
| 1808 | Qy | ----- | ----- | ArgGluArgGluLysSerIleLeuThr----- | 1816 |
| 9971 | Db | CTG | ACCGCGCAGCCCGCGGTGCGGAG | GTGAAATTGCGACGCGCACACGCCGACAC | 9912 |
| 1817 | Qy | ----- | SerThrThrThrValGluHisAla | -----ProlleTrpArgProGlyThrGluGln | 1833 |

| | | | |
|----|-------|--|-------|
| Qy | 942 | -----ProSerLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnL | 959 |
| Db | 12808 | CGGTGGCCACCGCGGCCACCACCGCGGCCACCGGGCCCCGGCGCGCGTGCACCGCGAAA | 12749 |
| Qy | 959 | ysProLeuAspLeu-----LysGlnLeuLysGlnArgAlaAlaAlaIleProProI | 976 |
| Db | 12748 | GGCCACTTAAGCTTACTGAAGGTGACGCTCTGAGCACCGAGGCGCGCGCGCGCGG | 12689 |
| Qy | 976 | leGlnValThr-----LysValHisGluProProArgGluAspAlaAap | 991 |
| Db | 12688 | TCCACCAACCCCGCGCGCACCACCGCGCCCACAGCCCGCGCGCACCCCGCGCGCGC | 12629 |
| Qy | 991 | roThrLysProAlaProProAlaProProProProGlnAsnLeu----- | 1005 |
| Db | 12628 | CGAACACCGCGCGCCCCCGCGGCCACCGCGCGCCACCGTGTCCTAAACCCCGCGTGC | 12569 |
| Qy | 1006 | -----GlnProGluSerAspAlaProGlnGlnProGlySerSerProA | 1020 |
| Db | 12568 | CGCGCGCACCGCGCACACCATCCACCGGCACCGCGCTGCGGTACAGCCACCGCG | 12509 |
| Qy | 1020 | rgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaGluAla- | 1039 |
| Db | 12508 | CGCACCGCGCGCACCGCGCACCGCGCGGCACCGCGCACCGCGCGCACCGCGGCAC | 12449 |
| Qy | 1040 | GlnLysLeuProGlyAspProProCystrThrSerGlyLeuProPheProValProPro | 1059 |
| Db | 12448 | CGATGTTCCCTCGAAGCGCGCGCGCGCGCACCGCGCACCA---CCGCGCGCGCG | 12392 |
| Qy | 1060 | ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro | 1079 |
| Db | 12391 | GCACCCCGCGTGGCCCCACAAACCGCGCGGTGCGCGCACACCGCGCGCGCGCGCAC | 12332 |
| Qy | 1080 | ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro | 1099 |
| Db | 12331 | CCGGCGCGCGTGTGGCG-----CCGGCGCGCGCGCGCGCGCGCGCGCGCG | 12296 |
| Qy | 1100 | ProThrIleSerAsnProProProLeuIleSerSerAlaLys-----HisProSer | 1116 |
| Db | 12295 | CGCGCGCGCGCAACCAACCGTTCGCGATCAGCCCGCGCGCACACCGTTCGCGCGCG | 12236 |
| Qy | 1117 | valLeuGluargGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValPro | 1136 |
| Db | 12235 | ATCC--CACGTGCTGCTGGTGGCGCGGTGGCCACCGTTCGCGTAGAG----- | 12189 |
| Qy | 1137 | TyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuPro----- | 1153 |
| Db | 12188 | -----CAACCGCGCGCGCGCGCGTCTTGTCCCGCGCGCGCGTGGC | 12147 |
| Qy | 1154 | -----LeuProMetAspProLysLysLeuAlaProPheSerGlyVal | 1167 |
| Db | 12146 | GCCATCACCGATCAGCGCGCGCCCCAAAAATCCGCTCGAAGCAGCGTGTGATCGCG | 12087 |
| Qy | 1168 | Lys-----ThrAlaGlnGluAlaSerValLeuArgGlyThrAla | 1173 |
| Db | 12086 | CAGTCTTGTTCACGCTCTGGCGCGCGAGAGTGTGACCGCTCGCAGCCGCATAGC | 12027 |
| Qy | 1174 | ProArgGlycInalaglyProProGluSerLeuGlyValPro----- | 1187 |
| Db | 12026 | CCCGCGCGCACCGTCAACGCTGATGAATCGGCGTGGGCGTGAACACCGCATTTGG | 11967 |
| Qy | 1188 | -----ThrAlaGlnGluAlaSerValLeuArgGlyThrAla | 1199 |
| Db | 11966 | CACCGCTGATACGCTGACCATGGCCCCCAACAGCAGCGATGTCGCGCGGAACCTC | 11907 |
| Qy | 1200 | Leu-----GlySerValProGlyGly-----serIleThrLysGlyIleProSerThr | 1215 |
| Db | 11906 | ATCAGCACCGCGGCCAACACCGCGGTGTTCGGCGCCAACCGCGCGGTTCGCGCG | 11847 |
| Qy | 1216 | ArgValProSerAspSerAlaIleThrTyrArgGly----- | 1227 |
| Db | 11846 | CAGTGTTCGACCGATCCCGCTAGATCCGAAGCGCGCGCGCAACACCTCGCGCGCAC | 11787 |

2090 lN-ProGlyProValIysLeuGlyGluAlaAlaHisLeuProHisLeu----- 2106
7590 GCGCTGGG-----TGGCGCATTTAGCCGGCTAGCAAACTCT 75628
2107 ---ArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGly 2125
75629 GTTCAACGTTGACCGCTCGCGGCCACGACGCTCGCGCGCGGACAGGCTCGCA 75688
2126 ValIysGlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGln 2145
75689 CGAACCGGTCATGAA----- 75703
2146 AspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSer 2165
75704 -----ACGTCGCCACTTGGCGGCTGACGCGTCTGAT 75733
2166 PheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp----- 2182
75734 ATTCTGGCGGTGGTGGCGGCCAACACCGCG---CAACGGCCACCGACACCTCGTCGCTG 75790
2183 -----LeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSer 2200
75791 ACGGGGCGACACTTTCGCCACCG-----CGCGCGCTGGGTGGCGGACGTGA 75841
2201 GluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAsp 2220
75842 TCGTCGAACCAATTTTCGCCAAATCCG----- 75868
2221 GlyIleGluPro-----ValSerProProGluGlyMetThrGluProGlyHisSerArg 2238
75869 ---TTGGCGCGGTGGTCAGCATCTCCGGCTCGCATACGACACATCTCGCTCCCA 75925
2239 SerAlaValTyrProLeuLeu-----TyrArgAspGlyGluGln-ThrGluProSe 2255
75926 GGTCAAGTCAGCGGTGTCGCCGGGTGCAAGGAATGTGTGGTATCCCGCGCATC 75985
2255 rArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLe 2275
75986 TACCATGGAGCGGAATCTTCGGGA-----TCCCAACT 76018
2275 uThrGluSer---AenSerAlaMetValLysSerLysGlnGluLeuAenLysLysLe 2294
76019 CCAACATCCCTGTTGACGTATCGTCAAAGGCAAAACCCCAAACTTTACGCGAACG 76078
2294 uAsnThrHisAsnArg-----AsnGluProGluTyrAsnI 2306
76079 AACTATCCACAGTCACCCCTCGATTTCGTGACACAGTGCACAGCGCCGACGATCGACGG 76138
2306 eSerGlnProGlyThr-----GluIlePheAsn----- 2315
76139 TCTAGCCCGCGCGGATATTGACAGTCTTCAGCGCGTTCGCGCGCGCGCGGAACTC 76198
2316 -----MetProAlaIleThrGlyThrGlyLeuMetThrTyr-----Ar 2328
76199 CGTTGCGCTCCCGCACCTCGCGGAGAGATAAGAAATGTTGCGCTATGTGGATGTCGG 76258
2328 gSerGlnAla-----ValGlnGluHisAlaSerThr-AsnMetG 2341
76259 GGCATACGCGCGAGCTCAACAGATTCTGTGGAGTTCGACGCGCGCTGACGCGTGGCGCG 76318
2341 lLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerP 2361
76319 GCGGTTCCGCGACCATCAGCATCGAGTCAAGATG---TGCTGAGGCGGATGG----- 76365
2361 roProLeuSerAlaAenAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaM 2381
76366 -----GCATTCCGCATACGAGGTGGATCTCATCTCGGTGAACGCGCA 76408
2381 etProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyG 2401
76409 TCCCGGGGACTTTCCTACCGCGCGGTTCGCCGCGACCGCATTTGCCCTACCTATGTT 76468
2401 lLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyL 2421

Db 76469 CGAGCGCTCGACATCGGTCGACCG-----CGAGGTT 76501
QY 2421 euAla-SerGlyAspArgProProSerValHisSerValHisSerGluGlyAspCysAsn 2440
Db 76502 GCGCCAGCGCGTGTGCGTAACCGCGCTTCGT-----CGT 76537
QY 2441 ArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThr 2460
Db 76538 CGACGTCAACT-----CGGCCAGCTGGCGGCTGCTTCG 76573
QY 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db 76574 GCTGTTGGGCTTCGACACACAGCGTGTGAGTGCAGCGCGCATGATCCGACGTCGGCGATAT 76633
QY 2481 ProProGlyLeuProAlaGlySerGly-----ProLeuAlaGlyPro--- 2494
Db 76634 CAGCGTGGCGGACGACGAATTCGTCTGACCGCGGACCGCGGCTGTGTAAGCGCGCGGC 76693
QY 2495 HisHisAlaTrpAspGluGluProLysPro 2504
Db 76694 AATCAACCATGCTGTTGTTCCATCCCA 76723
RESULT 78
AAI99683_12/c
Continuation (13 of 44) of AAI99683 from base 1200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 100001 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 200001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 400001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Db 73957 -----CCGCCGTTGCCGACCAAAACCGCCATCAC 73984
Qy 1412 luGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluG 1432
Db 73985 CGCCCTG----- 73992
Qy 1432 luLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly- 1449
Db 73993 -----CCGCCGAGCGCCCGTGTGGCAAAGCGGTGCATCGGCCCTCTGC 74041
Qy 1450 -----SerIleThrGlnGlyThrProLeu- 1457
Db 74042 CGCCGTTGCCCGCTGCCCGCCCTGGTGGGGGTGTGGCGGTGGCGCTTGGACCGG 74101
Qy 1458 -----LysTyrAspThrGlyAlaSerThrGlySerLysLysHisA 1472
Db 74102 GGGTGGAGCGCTTCCGCCCTGGCCCGCACCGCGACCGGATCA- 74151
Qy 1472 spValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspV 1492
Db 74152 -----CCGCCGTGGCCACCGCGCCACCTACACACCGCTTGACAC 74191
Qy 1492 alMetAlaAspAla-ArgAlaLeuGluArgAlaCysTyrClnuSerLeuLysSer- 1510
Db 74192 CGAGCGCGCGCGCGCGCGT-----GACCGCGGTGC-----CAGGAGTCCCGCGTCCCGC 74245
Qy 1511 -----ArgProGlyThrAlaSer-SerSerGlyGlySerIleAl 1523
Db 74246 CGGCTCGCGCTCACCGCAGCCCTGTGCTGCTGTGGCCCGAGGCGGCCAACGCGA 74305
Qy 1523 aArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSer-ProLeuThrT 1543
Db 74306 GACCGCGCGCGCGCGCTCGCCCGCGCTCCGCGAGCGCCACCGTTACCGCATTCCTCCG 74365
Qy 1543 yrClnuAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetA 1563
Db 74366 CGGGTGAAGCTCGCGCGCGCGCGAGCGACGCAATG-----AAGCCGATCTGCCAG 74416
Qy 1563 rgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspA 1583
Db 74417 CACCTCGGATCGCC-ATCG 74474
Qy 1583 rg-----LysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrV 1600
Db 74475 GGCCTGAGTTCCCGCTGCGCGCGCGCGCGCG-GTGCGCGCGCGCGCGCGCGCGCGCG 74533
Qy 1600 aProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerG 1620
Db 74534 CGGGCGCTTCGGATCCAAATCCGAGACCGCGCG-CCCGCGCGCGCGCGCGCGCGCGCG 74592
Qy 1620 lyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgG 1640
Db 74593 GCACCGCCATT-----ACCACCTACCGCGCTGCCACCCCTGCCACCGTTCCGCGCG 74646
Qy 1640 lyIle-ProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPro 1659
Db 74647 GTCTGTCCGCTGTGATAGCTCGCG------CCTTTGCGCGCGCGCGCG 74691
Qy 1660 ThrTyrProHisLeuTyr-----ProProTyrLeuIleArgGlyTyrPro 1674
Db 74692 CCGTTACCCGCTGTGGAGGTGGTGGCGCGCGCGCGCTGCCGCATCGCGCGTGGCG 74751
Qy 1675 AspThrAlaAlaLeuGluAenArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694
Db 74752 CCAGCGCTGCG- 74763
Qy 1695 MetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSer 1714
Db 74764 -----CCACCGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 74802
Qy 1715 ProArgGluSerSerLeuAlaLeuAenTyrAlaAlaGlyProArgGlyIleIleAspLeu 1734
Db 74803 CCATTGCCACG-----GTCCCGCGCGCGCACCAAG----- 74832

Qy 1735 SerGlnValProHisLeuProValLeu---ValProProThrProGlyThrProAlaThr 1753
Db 74833 ---TTGTGACCCACCGTCCCGGTAGCGCGGTTCGCGCGTACCGGAGTCTCCGCGTCA 74889
Qy 1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Db 74890 CCG-----CCGTACCGCCAGCCCGCGCGCGCGCGTGGTGGCGCGAGCGCG 74937
Qy 1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThr-ThrThrSerSe 1793
Db 74938 AGCAGACCGTGGCGCGCGCGCGCG-----CCGACCCCGCGCGTCCCG 74982
Qy 1793 rSerGluArgGluArgAspArgGluArgAspArgGluArgGluLysSe 1813
Db 74983 CCAGCCCCACCATTCGCGCGGTTCGCGCGGTGACCGTCAGCG- 75028
Qy 1813 rIleLeuThrSerThrThrValGluHisAlaProIleTyr-ArgProGlyThrGluG 1833
Db 75029 -----CCAAGTTGGTGGCGTGGGC----- 75047
Qy 1833 lnSerSerGlySerSerGlyGlyGlyGlyGlyGlyGlySerSerSerArgProAlaS 1853
Db 75048 --GCCGCTGGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 75105
Qy 1853 erHisSerHisAlaHisGlnHis-----SerProIleSerProArgThr-GlnAspAla 1870
Db 75106 TTGCGCGCGTTCGCCATCAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 75165
Qy 1871 LeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVal 1890
Db 75166 CCGCGCGCGCGCGCGCGCGT----- 75186
Qy 1891 GluProSerLysProThrValLeuArgSerThrSerThrSerProValArgProAla 1910
Db 75187 ---CCGATCCACCGGTGTGTGGCGGACGTA-----CCGTTGACACCGCGCG 75231
Qy 1911 AlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPro 1930
Db 75232 ATGCGCTTGCCTCGCGCGCGCGCGCG-----CCG 75264
Qy 1931 ThrLeuMetGluProValLeuLeuProLysGluAlaPro-ArgValAlaArgProGluAr 1950
Db 75265 ACACGAAACACCGCGCGTACCGCG-----GCCCGCGGTTCGCGCGCGCGCGCGCG 75318
Qy 1950 gProArgAlaAsp-ThrGlyHisAlaPheLeuAlaLysProAlaArgSerGlyLeuG 1970
Db 75319 GCCCGCCAAACCGCGCGCGCTCGTTGCCATACAGCACCGCGCGCGCGCGCGCGTGA 75378
Qy 1970 luProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerG 1990
Db 75379 --CCACCGCGCGCGCGGTGTACCGCGCGCGCTCCCGTTCGCGCGCGCGCGCGTACCG- 75435
Qy 1990 lyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProA 2010
Db 75436 -----ATTAGCGCGCGCGCGCGCGCGCGCGCGCTGTGTCTCT-----GGCGCGCGCG 75484
Qy 2010 spProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysP 2030
Db 75485 ACCCGCGGTGGCGCGGTTCGCGTACAGATGCG----- 75519
Qy 2030 roPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTy-Serp 2050
Db 75520 -----CCTGGCG 75538
Qy 2050 roGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyL 2070
Db 75539 CGGTCCCGGGGAGCGCGCGCGCG----- 75565
Qy 2070 euProLysHisLeuGluLeuLeuAspLysSerHisLeuGluGlyGluLeuArgProLysG 2090
Db 75566 TGCCG-----ATCAGCGCG-----CGTCCGACCA 75589

QY 790 SerArgAla----- 792
Db 72055 GCSCGGGGTGCAGACTGGCGGGTGGCGGATTTGGCCGATGATGCCAGGGGCGCG 72114
QY 793 -----ProIleGluProThrProAlaSerGluAlaThrGly 804
Db 72115 AGGTGTGGGTGGCGGGTGGCGGATTCACCGGTGCCACCTTACCGCCAGCCACCGGATCGGG 72174
QY 805 AlaProThrPro----- 818
Db 72175 AATAACCGCGGGTGGCGGGTGGCGGATTCACCGGTGCCACCTTACCGCCAGCCACCGGATCGGG 72234
QY 819 -----ProProValValProLysGluGluLysGluGluGluThrAlaAla 833
Db 72235 GTCTGAAGCGGTGGCGGATTCACCGGTGCCACCTTACCGCCAGCCACCGGATCGGG 72294
QY 834 AlaProProValGluGluGluGluGluGluLysProProAlaAlaGluGluLysAlaVal 853
Db 72295 AGGCGG-----TCGCGCGCGAA-GCCGCCCTTCGCGCGTC 72329
QY 854 AspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGly 873
Db 72330 ACCCGCGC---GCTAGCGTGTGGTGGCGCGCGCGCCCATGCGCGGTGGCGGC 72386
QY 874 ProAlaLysGly-----LysAspAlaGluAlaGluAlaThrAlaGluGlyAla--- 890
Db 72387 GGGGCCAAGCGGAGCGCAAGCGCGCACTGCGCGCGACCGCCAGCTTCGCGCGCGGC 72446
QY 891 -----LeuLysAlaGluLysGluGlySerGlyArgAlaThrAlaLys 907
Db 72447 GCGCGCGCGTGTGCG 72506
QY 908 SerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluAlaAsp 927
Db 72507 ACCACGGTGGCGC-----GGTGGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCG 72560
QY 928 GluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg-ProSerLeuLeuThrPr 947
Db 72561 GCCAGCGAGGTGCTGAAGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72620
QY 947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp---LeuLysGlnLe 966
Db 72621 GCGGTGGCG 72680
QY 966 LysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProAr 986
Db 72681 GCGGTGGCG 72737
QY 986 gGluAspAlaAlaProThrLysProAlaProAlaProPro----- 1000
Db 72738 GCGCGCGGTGCG 72797
QY 1001 -----ProProGlnAsnLeuGlnProGln 1008
Db 72798 GTCGAATGTCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72857
QY 1008 uSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPr 1028
Db 72858 GCG 72917
QY 1028 oProAlaAspLysGluAlaPheAlaGluAlaGlnLysLeuProGlyAspPro----- 1046
Db 72918 GCGTACGCGC-----CGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72944
QY 1047 -----ProCysTrpThrSerGlyLeuPro---PheProValProProArgGluValI 1063
Db 72945 GCGCGTGGCGGTAATGGCTCCGGGATTCGCGTCACTACCGGTGGCGCGCGCGCGCGCG 73004
QY 1063 eLysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaProProGlyHisPr 1083
Db 73005 GCG 73064
QY 1083 oLeu-----ProLeuGlyLeuHisAspThrAlaArgPr 1094

Db 73065 GTACCGCCCATGCCCCCTGTGGCATTTGCCGCGGTGGCGCGCGCGCGCGCGCGCG 73124
QY 1094 oVal-----LeuProArgProProThrIleSerAsnProPro-----ProLeuI 1109
Db 73125 GCTCGCGCGAGGTGGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCG 73184
QY 1109 eSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMe 1129
Db 73185 GCTGAGGCAATGGTGGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCG 73244
QY 1129 tSerValGlnLeuHisValProTySerSerGluHisAla-----LysAlaProVa 1145
Db 73245 GCTGCTTGGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCG 73304
QY 1145 lGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLysLeuAlaProPheSe 1165
Db 73305 GGGCGCGCGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCG 73353
QY 1165 rGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGl 1185
Db 73354 ---GTGAACACGTCGCTTAGCGCGCTG-----CGCGCGTCAAC-GG 73393
QY 1185 yValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySer----- 1202
Db 73394 GGTCCCGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCG 73453
QY 1203 ---ValProGlyGlySerIleThrLysGlyIleProSerThrArgVal-----Pr 1218
Db 73454 CCG 1227
QY 1218 oSerAspSerAlaIleThrTyArgGly-----ProSer-----ArgLeuAs 1256
Db 73514 CCG 73573
QY 1228 ---SerIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrIleThrAr 1245
Db 73574 CCGCGCGCGCGTACCGCAACTCCCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCG 73633
QY 1245 gIleIleGlyGluAspSer-----ProSer-----ArgLeuAs 1256
Db 73634 CTGTGGTGGCG 73693
QY 1256 pArgGlyArgGluAspSerLeuProLysGlyHisValIleTyLysGlyLysGlyHi 1276
Db 73694 CCGCGGTACCGCAACTCCCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73753
QY 1276 sValLeuSerTyLysGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSe 1296
Db 73754 CGTTGACACGTCGTCGCGGACCCCTCC-----AGTCCCGCGCGCGCGCGCGCGCG 73801
QY 1296 rSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyAspMetMetGluGl 1316
Db 73802 TGCG 73832
QY 1316 yArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAla-IleP 1336
Db 73833 -CGCCCG 73879
QY 1336 roProGluArgHisSerProHisHisLeuLysGluGlnHisIleArgGlySerIleT 1356
Db 73880 CGCATTCACG----- 73890
QY 1356 hrGlnGlyIleProArgSerTyValGluAlaGlnGluAspTyLysArgArgGluAlaL 1376
Db 73891 --CGGGGTGGCGAGC----- 73905
QY 1376 yLeuLeuLysArgGluGlyThrPro-----ProProProProSerArgA 1392
Db 73906 ---CTGCCAGACCG 73956
QY 1392 spLeuThrGluAlaTyLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisG 1412

QY 2504 roLeuLeuCysSerGlnTyrGluThrLeuSerAspSer 2516
Db 61134 CTCGA-----TCGATCACCAACGTTGGTCCCGTCC 61103

RESULT 76

AAI99683_06
Continuation (7 of 44) of AAI99683 from base 600001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

| Fragment Name | Begin | End |
|---------------|---------|---------|
| AAI99683_00 | 1 | 110000 |
| AAI99683_01 | 100001 | 210000 |
| AAI99683_02 | 200001 | 310000 |
| AAI99683_03 | 300001 | 410000 |
| AAI99683_04 | 400001 | 510000 |
| AAI99683_05 | 500001 | 610000 |
| AAI99683_06 | 600001 | 710000 |
| AAI99683_07 | 700001 | 810000 |
| AAI99683_08 | 800001 | 910000 |
| AAI99683_09 | 900001 | 1010000 |
| AAI99683_10 | 1000001 | 1110000 |
| AAI99683_11 | 1100001 | 1210000 |
| AAI99683_12 | 1200001 | 1310000 |
| AAI99683_13 | 1300001 | 1410000 |
| AAI99683_14 | 1400001 | 1510000 |
| AAI99683_15 | 1500001 | 1610000 |
| AAI99683_16 | 1600001 | 1710000 |
| AAI99683_17 | 1700001 | 1810000 |
| AAI99683_18 | 1800001 | 1910000 |
| AAI99683_19 | 1900001 | 2010000 |
| AAI99683_20 | 2000001 | 2110000 |
| AAI99683_21 | 2100001 | 2210000 |
| AAI99683_22 | 2200001 | 2310000 |
| AAI99683_23 | 2300001 | 2410000 |
| AAI99683_24 | 2400001 | 2510000 |
| AAI99683_25 | 2500001 | 2610000 |
| AAI99683_26 | 2600001 | 2710000 |
| AAI99683_27 | 2700001 | 2810000 |
| AAI99683_28 | 2800001 | 2910000 |
| AAI99683_29 | 2900001 | 3010000 |
| AAI99683_30 | 3000001 | 3110000 |
| AAI99683_31 | 3100001 | 3210000 |
| AAI99683_32 | 3200001 | 3310000 |
| AAI99683_33 | 3300001 | 3410000 |
| AAI99683_34 | 3400001 | 3510000 |
| AAI99683_35 | 3500001 | 3610000 |
| AAI99683_36 | 3600001 | 3710000 |
| AAI99683_37 | 3700001 | 3810000 |
| AAI99683_38 | 3800001 | 3910000 |
| AAI99683_39 | 3900001 | 4010000 |
| AAI99683_40 | 4000001 | 4110000 |
| AAI99683_41 | 4100001 | 4210000 |
| AAI99683_42 | 4200001 | 4310000 |
| AAI99683_43 | 4300001 | 4403765 |

Alignment Scores:

| Pred. No.: | Length: |
|------------------------|-----------------|
| Score: | 110000 |
| Percent Similarity: | 614.50 |
| Best Local Similarity: | 32.06% |
| Query Match: | 22.54% |
| DB: | 4 |
| | Mismatches: 905 |
| | Indels: 108 |
| | Gaps: |

US-09-522-753-5 (1-2517) x AAI99683_06 (1-110000)

QY 355 GlyClnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGlu 374
Db 72087 GGGATAGCGCCTCTGACCTGACCGACAGACACCGGTGTACGGAATATGACGTTTCGCAA 72146
QY 375 IleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaVal 394
Db 72147 CTGGCGGATCATTTG-----CTGCGCTCGTGGCGATC 72179

QY 395 Ile-----ProProMetLeuTyrAspAlaAspGlnGlnArg 406
Db 72180 ATCGCGCGTGGCGCGCGCGAGCTGGCGCCCGCGATGTGGAGCGCCA----- 72230
QY 407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArg 426
Db 72231 -----CTGGAAACCCAGGTGGCGGACGCGCGCAGGCC----- 72263
QY 427 GlnValMetAsnMetTyr-----SerGlu 434
Db 72264 --GTGATGGAACCTGGCGCGCGTGGTGGCGGCGACCGTGGAGCTGAACTCGAAC 72320
QY 435 GlnGluLysGluThrPheArgGlu-----LysPheMetGlnHisProLys 449
Db 72321 CAGTGCCTCGACGCGTGGCGGTCGCGATCCTGTCGTAGAAATTTCTGGTCCACGCTGG 72380
QY 450 AsnPheGlyLeuIle-----AlaSerPheLeuGluArgLysThrValAlaCysVal 467
Db 72381 GATTCGCGATTGCCACCGGTTCTCAGGTGATCGCGTCCGAGCGGTGTGGAGTACGTA 72440
QY 468 LeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr 487
Db 72441 CTGG--CGGTGGCGCGCAAGGTCTACCCCGCAACCCCTAACTCCGCGGGTTCGCCG 72498
QY 488 ArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 507
Db 72499 CGCGCGCGCGTGGTTCCTTTGCCCGCAGTCTCGATCGCTCATCGCTTACCAGGCC 72558
QY 508 GlnGlnGlnProMet--ProArg--SerSerGlnGluGluLysAspGluLysGluLysGluL 527
Db 72559 GCCAGCGCGCGCGCGCGCGTGTCCGCCACCTAACGAAAGGATGATCATGCCCAAGAGA 72618
QY 527 YsgluAlaGluLysGluGluGluProGluValGluAsnAspLysGluAspLeuLeuL 547
Db 72619 AGCGAATACAGGCAAGCAGCGCGAACTGGTGCAGCTTCAGACCCAGCATCGTCCGCC 72678
QY 547 YsgluLys----- 549
Db 72679 GCCAAAAGTTCTACACATCGTTTGGTGGGTTCAGCAGCAACCCGGTCCCGGA 72738
QY 550 -----Thra 551
Db 72739 GCGGTGGGTCTATTCCATGCGCAACCGGAAAGCCGTGGCGCGCATCGCACCG 72798
QY 551 spAsp-----ThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysG 568
Db 72799 ATGCCCGCGGTGCACCGGAGGGATGCCCGCATCTGGAACACCTATATCGCGGTGGAC 72858
QY 568 lyArgLysThrAlaAsnSerGlnGlyArgLysGlyArgLleThrArgSerMetAlaA 588
Db 72859 GACGTGATCGGTGGTGGACAGGTGTGTCGCGGGGCGG----- 72899
QY 588 snGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGlu--LeuAlaSerMet 607
Db 72900 -----GCAGGTGATGTCGCGCGCTTC 72921
QY 608 GluLeuAsnGluSerSerArgTyrThrGluGluMetGluThr---AlaLysLysGly 626
Db 72922 GACATCGCGATCGCGCGCGATGTCGTCATCCGATCCGACCGCGGTGCGGTGGCG 72981
QY 627 LeuLeuGluHisGlyArgAsnTyrSerAlaIleAlaArgMetValGlySerLysThrVal 646
Db 72982 CTA-----TGGCAGGCC---AATCGGCACATCGGAGCGAGCTTGGTC 73020
QY 647 SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeu 666
Db 73021 AACGACCGCGCGACGCTCATCTCGAACGAACTGCTCAGGACCAAGCGCGATTGGCGCTA 73080
QY 667 GlnGlnHisLysLeu-----LysMetGluLysGluArgAsnAla 679
Db 73081 CGGTTCTACGAGCGCTGGTGGCTTCCACCTCAGCATCGAGTATGATAGTGGCGGCGAG 73140
QY 680 ArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProVal 699

| | | | | | |
|----|-------|--|----|-------|---|
| Db | 63040 | CCGCGACCAAAACCGACCGCCACCATGATCAG | Db | 62133 | CGCCAGATCAAGACCCCGCCACATCGCGCGCGATCTCCAGCATCGAATCCCGAG |
| Qy | 1990 | valGluProSerLysProThrValLeuArgSerThrSerSerProValArgPro | Qy | 2206 | erProGlu |
| Db | 62998 | ACGACACCCAGACCGCGACGCGCGGTACGACACCC | Db | 62073 | CACCACATCCGCGCGACCCCGACCGACTCCCAACCGGG |
| Qy | 1910 | AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyr | Qy | 2223 | luProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrP |
| Db | 62950 | GTTCGGGAACCTCACCGCG | Db | 62032 | -----CAAACCCACTCGCGCGCAACACCCCGCTCGCCACATCGTGT |
| Qy | 1930 | ProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGlu | Qy | 2243 | roLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProG |
| Db | 62926 | CCAAACAGGATGCCCG | Db | 61984 | -----GATCCACACCGCTCCGCGCCACGGAACACACCTCCCG |
| Qy | 1950 | ArgProArg | Qy | 2263 | lyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetV |
| Db | 62875 | CCCAGAACGGCGCTCAACAGAACCGCTCCCGCTGGAGCGATAGCTCGGCACTCCA | Db | 61944 | CA---ACCACCGCGCTCCAGCCCAACCGCCCTCCACAGGC |
| Qy | 1964 | ProAlaArgSerGlyLeuGluProAlaSerSerProSer | Qy | 2283 | allYsSerLysLysGlnGlnLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluProG |
| Db | 62815 | CCAGCCAGGACACGATCACCCGGAACAACACCGACGATCCACAGCAACACGCGG | Db | 61903 | -----CGCACACCTCATCAACGACGCGCG |
| Qy | 1977 | -----LysGlySerGluProArgProLeuValPro | Qy | 2303 | luTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly |
| Db | 62755 | TATGACAGACGACCATCCCTCCAGAACGCGGACTCTCAGHACGCTCAGCGCCAGCG | Db | 61878 | AA-----ACCGGGAACCGCTCATCAAAACACAGCCCATACCCCAACTG |
| Qy | 1987 | -----ProValSerGlyHisAlaThrIleAlaArgThrProAlaLys | Qy | 2322 | -----ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisA |
| Db | 62695 | AGGACACAGTACCGTTCCGGGCGCATGCTATCGGTTACGTCGTCAGGGTG-TGCTGG | Db | 61830 | ACTGCCCTCAGCACTGACAGCCACACACACCGCTCGCAGCAACCCGCTCTCAGC |
| Qy | 2001 | -----AsnLeuAlaProHisAlaSerProAspProProAlaProAla | Qy | 2336 | laSerThrAsnMet |
| Db | 62636 | GCGGAGTAGTCAGTACGGGCGCGGCGGAGCTCGAACAACACGCGCTCGGGAAGCC | Db | 61770 | GGACCGCGTCACAGCGGGTGTCTCTCGCCAGCGGCAACACAGCAAGCCCGCAGAGG |
| Qy | 2017 | SerAlaSerApproHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeu | Qy | 2353 | ysTyrAspGlnTrpGluGluSerProProLeuSerAlaAlaAsnAlaPheAsnProLeuAsnA |
| Db | 62576 | ACTGGCGATGCGGGCTGGAAGACACTGTCTGGCGTACATGGCGTCCGAG | Db | 61710 | TTCAGCCCTCTCAGGCCCGCCACACACCGCTCTGCTGAACAGCGCAGCGCTCTCAG |
| Qy | 2037 | GluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProVal | Qy | 2373 | laSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisT |
| Db | 62522 | -----TACTCCGGGAGCGATCTCC | Db | 61650 | CAGCGACCGC---CCACATCTCTGCACACCGTGTGATCGCGCGCCACCGCATCAGC |
| Qy | 2057 | SerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeu | Qy | 2393 | hrLeuThrSerProGlyGlyGlyLysAlaLysValSer |
| Db | 62501 | TCAGCCGCTCCAGACCGAGAC-GTGTCTGATCAA | Db | 61593 | CAGCGTAGTGTCTGTCTGCGCAGCGGCTGCGGGCTGCGGGCGATACACCCACCGCAC |
| Qy | 2075 | GluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProVal | Qy | 2409 | roSerSerArgLysAlaLysSer |
| Db | 62451 | CGTGAATTTGACTCTCGTATCGTTCGTGTAACCTCCGCGAGCATCGGCTCCATCAACGC | Db | 61533 | CAAGCGCGCTCAGTCTCAGTCTCGGGAGCTCACCACCTGACCGATACCGAGTCTCTCCC |
| Qy | 2095 | LysLeuGlyGly-----GluAlaAlaHisLeu | Qy | 2421 | euAla--SerGly |
| Db | 62391 | CGAGTGAAGGATGACTGACGCTCAGCGCTTCTTACGCCACCGCTCCCGCCACAC | Db | 61473 | CAGCCATCGGGCACAGACTCCACGATCAGTGTGCGATTCGTTCCCGAAGCCCGAAGCA |
| Qy | 2110 | ProGluSerGlnPro-Ser-Ser-SerProLeuGlnThrAlaProGlyValLysGlyHis | Qy | 2425 | -AspArgPro |
| Db | 62331 | CCC---AGCAATCCGATCCACCTCACCA | Db | 61413 | CGACACACCGCGCGACCGGACGACCTCCACTCGGCCATGGCACCGACTCAACCGC |
| Qy | 2129 | sgIna-gValValThrLeuAlaGlnHisIleSerGluValIleThr-GlnAspTyrThrA | Qy | 2439 | sAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSer |
| Db | 62283 | ATCAGGTGTGTGACCGCGCC | Db | 61353 | CAACCGCACCGCACCCGAAATCCCACTCCACATGCGCGGACGCGCATCCCATGCAAACT |
| Qy | 2149 | rgHisHisProGlnGlnLeuSerAlaProLeuPro-----AlaProLeuTyrSerPheP | Qy | 2456 | -AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMe |
| Db | 62238 | GTCCGGCGGCGAGTCCCGGGCGTGGCTGCACCGCGGCACATCGCCCCCG | Db | 61293 | CGCGGCGCACACCA |
| Qy | 2167 | roGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProP | Qy | 2476 | tAlaSerProProProGlyLeu-ProAla |
| Db | 62187 | CTCAGGAGCCACCCATCAACAGCGCGCGCGCCCGCCACCCACCGCAGGCAT | Db | 61195 | GATGACACCCGCAACACCGCGCAGCGCGCTGGGTATGACCAATATTCGACTTCACCGACCC |
| Qy | 2187 | roProAspHisGlyAlaProAlaArgGlySerProHisSer---GluGlyGlyLysArgS | Qy | 2504 | GlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysP |
| Db | 2206 | | Db | 61194 | CAACCAACAGGAGCATCCACCGGACGCGCCCTGCCCATACGTGCGCAGCAGAGCTCGGC |

Qy 1244 ThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeu 1263
Db 64798 GCGCGACCATCGGG-----TGTCGCCACGG----- 64772
Qy 1264 ProIleGlyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGly 1283
Db 64771 CCGTCGGGGCGAGGACCGTGGAGGACGCCCATCACGC----- 64733
Qy 1284 MetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGlu 1303
Db 64732 -----GGTCGCCCGACGAGTGGGTGCATCGG 64703
Qy 1304 ThrAlaAlaProIysArgThrTyrAspMetMetGlu---GlyArgValGlyArgAlaIle 1322
Db 64702 GGCCTGTCTCCGTCCACACACCGCGCGCTTCGCGCCCGACGCGCATCGC----- 64644
Qy 1323 Ser-----SerAlaSerIleGluGlyLeuMetClyArgAlaIleProProGluArg 1339
Db 64643 ATGCCGAGGCTGACCGAGCGTCCGGAAGTTGATCCCGCGCGGTGATGCCGAGCGCG 64584
Qy 1340 HisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIle 1359
Db 64583 ACCTGGCGGGGGCC-----AGGGTTCCATGACTCCGGGGTAC 64545
Qy 1360 ProArgSerTyrValGluAlaGlnGluAspTyrLeu-----ArgArgGluAlaLys 1376
Db 64544 GCGACCGCGCGACGCTCTCCAGCGTCTCGCGCTTCCTTGTCCAGCGCCAGGCACGG 64485
Qy 1377 LeuLeuLysArgGluGlyThrProProProProProProProProProProProPro 1396
Db 64484 TCGGTG-----GGCGCCACGACGAGTCCCGCGCGGTGGCGGA----- 64445
Qy 1396 aTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAl 1416
Db 64444 -----CCAGCCCGGAACGA-----GC 64428
Qy 1416 aThrValLysGluAlaGlyArg---SerIleHisGluIleProArgGluLeuArgHis 1435
Db 64427 AGTTGTCGTCCGCGCGGCGAGTTGGGGCTCGTCGAGTTCCTCGCGCGGTGA-CCGAG 64369
Qy 1435 s-----ThrProGluLeuProLeuAlaPr 1443
Db 64368 GTGGCGGTGGGAGTCTCGCGCGTTCAGTCCGGGTCCAGTCCGTCGTCATTTG---CC 64312
Qy 1443 oArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAl 1463
Db 64311 GTCTCCGTCCACGTCCACGTCCAGCTCCAGCTCCACGTCCATGTCTCCAGTCGAC 64252
Qy 1463 aSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgTh 1483
Db 64251 GTCCACGTCC----- 64242
Qy 1483 rPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCy 1503
Db 64241 -----ATGTCACGTCCAGCAGGATGAAGCGCGCCCGG 64210
Qy 1503 sTyrGluGlu---SerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIl 1522
Db 64209 GTTCTCGGATGACATCGGACCGAGCCCATACGGCTGCGGGGACGCGTCCCATC 64150
Qy 1522 eAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuTh 1542
Db 64149 ACCGTCCCGTCGCG-----CCGGTCGCGCACGGCGCGCGGTGACACGAGGAG 64099
Qy 1542 rTyrGlu-----AspHisGlyAla-PropheAlaGlyHisLeuProArgGlySerProv 1560
Db 64098 CCGGAGTCGGCGGACGTCCGCGCGCGGCA-----GGAACTCCTTG 64060
Qy 1560 alThrMetArgGluProProThrProArg---LeuGlnGluGlySerLeuSerSerLysA 1579
Db 64059 TACGAGTGACAAACCCCGTCCACGACCGACCGTCAAGACCTTCACGCGCGGTCAA 64000

Qy 1579 laSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerT 1599
Db 63999 GCGCGCAAGCGCGCGC-----CCGCGATGACCCGAGCGGACGACCTGC 63952
Qy 1599 hrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValS 1619
Db 63951 CMAGCGCGTTCACCAACCGCGCGC-----ACACCATCATCAG-----C 63913
Qy 1619 erGly-ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db 63912 CAGGCTCCGACACACCGTCCGAGAGCACACCCAGTC----- 63877
Qy 1639 ArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn 1658
Db 63876 ---GGCATCCCGTCCGTCATCCGTCGACACATCCGTCGCGACAGCGGCGCGT 63820
Qy 1659 ProThrTyrProHisLeuTyrProProTyrLeuIleArgGly-----TyrProAspThr 1676
Db 63819 CCAGTCCCGAGAACAGACCGTCTGTGTGACCCCGGACGCGCTGCAGTTGCCGATGTC 63760
Qy 1677 AlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyr-----IleThrSer 1692
Db 63759 GCGCGGACGCAAGCTCACCGACCCCGACGTCGAGTCCGTCGCGTCCGTCGCGCAC 63700
Qy 1693 GlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnAlaAspMetLeuArgGly 1712
Db 63699 CAGTACCTGTACTTCCGTTCCGCTTCGCTGCTG----- 63667
Qy 1713 LeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly----- 1727
Db 63666 CTCGCGCGGCGAGAGACG-----TACGCGTACGTTGGCGGCTCCCGGAGC 63622
Qy 1728 ProArgGly---IleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
Db 63621 CCGCAGGAGACCTCGTTCAGTGAAGGACGACGCTTCCCTCGTCTGCTGCTGCAC 63562
Qy 1747 ThrPro----- 1748
Db 63561 CTCGCGCGGCGACGTCGAGGACGCGCGGTGGAGGTGGGCTCGAGGAGGCGGGTG 63502
Qy 1749 -----GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPro 1762
Db 63501 GATGCGGTAGCGCTCATGGGACCCCGCGCTTCGGGACAGACACCTCGGCGAGGAGTC 63442
Qy 1763 ThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyPro 1782
Db 63441 CTGACCGTCCGCCA-----CACGCGACGACGCGCTTCGACACCCCG-GGCGGT 63395
Qy 1783 ThrHisLeuThrLysProThrThrThrSerSerSer----- 1794
Db 63394 AGCATAGCCCGCTCCGCGACGCGCTCGTAGAAGCCCGCAGGTCCACCGGACGCGC 63335
Qy 1794 ----- 1794
Db 63334 CCGCGCGGCGCACGCGCGCGAGTCCGTCGTTGGAGCGTGTGCTGGGGGTGA 63275
Qy 1795 -----GluArgGluArgAspArgGluArgAspArgGluArgGlu 1811
Db 63274 GGTGCGCGGTGCGTGGGACATCCAGAGTCTGCGCGCTCAGG---CGTCCGCTCAG 63218
Qy 1812 LysSerIleLeuThrSerThrThrValGluHisAlaProIleTrpArgProGlyThr 1831
Db 63217 GCGCGGTATAGATCTGACACCGCGCGCGCTCTCCTCAGCCGCTCTCGACACCACT 63158
Qy 1832 GluGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlyGlySerSerSerArgPro 1851
Db 63157 GAACCGGACGCGCGCTCAGGACCCACCAAGCGCGCTGCAACATCACTCTCTCA 63098
Qy 1852 AlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg-----ThrGlnAsp 1869
Db 63097 CCGTCACACACCGACTCAT---CACCGCGCGCAACGCGCCACTCCACCAAGCGACAC 63041
Qy 1870 AlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla 1889


```
Qy 159 GluLeuGluLeuValPro-----ProArgLeuSerLysGluGlu 171
Db : : : : :
68813 CAGTCCCTCCAGCAAGCGGCACACCTCATCAAGCGCAGCGCAACACCGGAAACCG 68754
Qy 172 LeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSer 191
Db : : : : :
68753 CTCATCAAAACGACGACCCATACCGCAACACTGACTGCCCTGACCACTGAACACGACAC 68694
Qy 192 LysLeuLysLysGlnGln----- 199
Db : : : : :
68693 CACATCACCGCTCGCAACACCGCGCTTCTCAGGAGCGCGTACAGCCGCGGTCTC 68634
Qy 200 ---LeuGluGluAlaAlaLysProProGluProGluLysProValSerPro----- 216
Db : : : : :
68633 CTCACACGCGCCAAACGACGACCGCGCGCAGCAGCTTACGCCCTCTCACGCCCGCCAC 68574
Qy 217 -----ProIleGluSerLysHisArgSerLeuValGlnIleLeuIleThrAspGluAsn 234
Db : : : : :
68573 CACCGCGCGCTGCTCGAACAGCGCAGCGCAGCCGCAAGACGACGCGCTCGGTGAG 68514
Qy 235 ArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu 254
Db : : : : :
68513 CGCGCTCGGAGCGCTACCGCGCAAGTCCGCAGAC-CCAGGGGCTGAT----- 68464
Qy 255 ProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGln 274
Db : : : : :
68463 -----CACGACGCGCTCGGACTACGGCGCGCACACCCACGCGCACCGAGCGCT 68410
Qy 275 AlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLys----- 292
Db : : : : :
68409 CGAAGCGCGCGGGTTCGCGCGCTTCGCGCGCGTTCGCGCGCGTTCGCGCGCGCGCG 68350
Qy 293 -----GlnTyrLysGlnLysPheCysGlnArg----- 301
Db : : : : :
68349 GGGCTGTTCCAGATCAGTGGCGGTGG-----TGCGCTGTATGCCGAAGAGGAGACCC 68293
Qy 302 -----TyrAspGlnLeuMetGluAlaLeu----- 309
Db : : : : :
68292 CCGCGCGCGCGCGTGGCGCGTGGCCACGCGCAGCGATCGTGAGCAACCGCACTC 68233
Qy 310 GluLysLysValGlu-ArgIleGluAsnPro----- 320
Db : : : : :
68232 CCGCGCGCTCCAGTCCAGCTGGCGCGTGGCGCGTTCGATGTGCGAGGCGCGCAGCT 68173
Qy 321 -----ArgArgArgAlaLys----- 325
Db : : : : :
68172 CGCGTGGCGCATCGCCATCACCATTGTATGACGCTCGCAGCGCGCGCGCTGCG 68113
Qy 325 sGluSerLysValArgGluTyr---TyrGluLysGlnPhePro----- 338
Db : : : : :
68112 TGTGGCGATGTTTCGACTTCACAGAGCGGACGCGCGCGTGGCGCGCGGTCTCT 68053
Qy 339 -GluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgG 358
Db : : : : :
68052 TTCGTACGTCCGAGAGTGCCTGCC-----CTCGATCGCGTCCGCGCGGTGG 68002
Qy 358 ySerGlyLeu-----SerMetSerAlaAlaArgSerGluHisGluValSerG 374
Db : : : : :
68001 TGGCGTGGCGTGTGCTCCACGCGGTGCGAGCGGTCCGGGTGACCGCGCGGTGTGTCA 67942
Qy 374 uIleIleAspGlyLeuSerGlnGlnAsnLeuGluLysGlnMetArgGlnLeuAlaVa 394
Db : : : : :
67941 G-----GGCTGGCGGATCAGCGGTTCCTGGCGGAGCTGCTTCGGCGCGGTGAGC 67891
Qy 394 lIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnG 414
Db : : : : :
67890 CGTT-----GCTCGCGC 67879
Qy 414 yLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetIrpSerG 434
Db : : : : :
67878 CGTCTGTTGACGCGCGATCCCGGTACGAGCGCGAGTACTTGTATGTCGTT---GCGGC 67822
```

| | | | |
|------------|---|--|------|
| Db | 6288 | CCTGGCGCTGGCGGGCGGACCGCCACACCCCTCGACGCGCTCTACGACGCGCTCGCG | 6347 |
| Qy | 2420 | Y-----LeuAlaSerGly | 2424 |
| Db | 6348 | AGCTCGGCTACGATACGGTCCGCGCTTCAGGGCTGACGGGCTGTGGCGGACGGCG | 6407 |
| Qy | 2425 | -----AspArgProProSerValHisSerGluGlyAspCysAsnAr | 2441 |
| Db | 6408 | CCGACAGCTCGCGAGATCGGCTGCCGCGGCGACGACGAGAGCGGGGCTCTTCG | 6467 |
| Qy | 2441 | gArgThrProLeuThrAsnArg-----ValTrpGluAspArgProSerSerHisGlyse | 2459 |
| Db | 6468 | GCCTACACCGCGCTGCTGACGCGCGCTCCACCGAGTCGCTCGAGG-SCAACTCA | 6526 |
| Qy | 2459 | rThrProPheProTyrAsnProLeuLeuMetArgLeuGlnAlaGlyValMetAlaSerPr | 2479 |
| Db | 6527 | GCTCGCGGTGCTGTGACGCGGATA----- | 6551 |
| Qy | 2479 | oProProProGlyLeuProAlaGlySerGly----- | 2489 |
| Db | 6552 | -----CCGACGCGACCGACCGGATCCGGCTGCGTTCGGTGGCGGGGTGACCCCTC | 6604 |
| Qy | 2490 | -----ProLeuAlaGlyProHisHisAlaTrpAspGluGluPro | 2502 |
| Db | 6605 | CACGCCGAAGGGCGCACCGCGCTCCGCGTACGGATCACACCCA | 6647 |
| RESULT 75 | | | |
| ABX04971/c | | | |
| ID | ABX04971 | standard; DNA; 103599 BP. | |
| AC | ABX04971; | | |
| XX | 16-JAN-2003 | (first entry) | |
| DT | | | |
| XX | S. cinamonensis monensin type I polyketide synthase gene cluster. | | |
| DE | Monensin; gene; cluster; polyketide synthase; antibiotic; de; | | |
| KW | antihelminthic; insecticide; immunosuppressant; antifungal; | | |
| KW | antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H; | | |
| KW | mon RI; mon RII; mon T; mon AIX; mon AX. | | |
| KX | Streptomyces cinamonensis. | | |
| OS | WO200168867-A1. | | |
| XX | 20-SEP-2001. | | |
| XX | 30-MAY-2000; 2000WO-GB002072. | | |
| PF | 28-MAY-1999; 99GB-00012563. | | |
| PR | (BIOT-) BIOTICA TECHNOLOGY LTD. | | |
| XX | Leadlay PF, Staunton J, O'liynyk M; | | |
| XX | WPI; 2001-611393/70. | | |
| DR | P-PSDB; ABG99854, ABG99855, ABG99856, ABG99857, ABG99858, ABG99859, | | |
| DR | ABG99860, ABG99861, ABG99862, ABG99863, ABG99864, ABG99865, ABG99866, | | |
| DR | ABG99867, ABG99868, ABG99869, ABG99870, ABG99871, ABG99872, ABG99873, | | |
| DR | ABG99874, ABG99875, ABG99876, ABG99877, ABG99878, ABG99879, ABG99880, | | |
| DR | ABG99881, ABG99882, ABG99883, ABG99884, ABG99885, ABG99886, ABG99887. | | |
| XX | New DNA sequence encoding polyketide synthase, useful for the production | | |
| PT | of polyketides such as antibiotic monensin. | | |
| XX | Claim 1; Page 116-195; 212pp; English. | | |
| XX | The invention relates to a DNA sequence which is a fully defined sequence | | |
| CC | of 103551 base pairs appearing as ABX04971, or its variant, that it is | | |
| CC | not a sequence encoding all or part amino acids 1-920 encoded by mon AI | | |
| CC | as given in the specification. The DNA is the S. cinamonensis polyketide | | |
| CC | antibiotic monensin biosynthetic gene cluster. Also included are a | | |

| | |
|----|---|
| CC | recombinant cloning or expression vector comprising the gene cluster, a |
| CC | transformed host cell which has been transformed to contain the gene |
| CC | cluster (and is capable of expressing a corresponding polypeptide), a |
| CC | hybridization probe derived from the gene cluster (for identification and |
| CC | isolation of the same or analogous gene cluster, e.g. one which binds |
| CC | specifically to a region of the monensin gene cluster selected from mon |
| CC | BI, mon BII, mon CI, mon CII, mon H, mon RII, mon T, mon AIX and |
| CC | mon AX), the use of the mon RI gene or variant and a monensin promoter to |
| CC | control expression of a heterologous gene in Streptomyces cinamonensis, |
| CC | a polypeptide encoded by a portion of the monensin gene cluster |
| CC | (preferably comprising mon BI, mon BII, mon AIX or mon AX or their |
| CC | mutants, alleles or variants), an epoxidase encoded by mon CI, a |
| CC | cyclase encoded by mon CII, producing S. cinamonensis capable of |
| CC | enhanced levels of production of monensin comprising multiple copies |
| CC | of the mon RI gene, S. cinamonensis containing multiple copies |
| CC | of the mon RI gene and/or its variants, expressing a gene heterologous to |
| CC | S. cinamonensis comprising transforming S. cinamonensis with DNA |
| CC | encoding a heterologous gene and expressing the gene under control of the |
| CC | activator gene mon RI or actII/orf4 and 13-propyl erythromycin A. The |
| CC | processes and materials (enzyme systems, nucleic acids and vectors) are |
| CC | useful for preparing polyketides by recombinant synthesis. The |
| CC | polyketides are useful as insecticides, antibiotics, antihelmintics, |
| CC | antifungals, antibacterials or other pharmaceuticals. In particular the |
| CC | gene is useful for the production of monensin, an antibiotic polyether |
| CC | polyketide. The present sequence represents the monensin gene cluster |
| XX | Sequence 103599 BP; 13980 A; 37023 C; 37799 G; 14795 T; 0 U; 2 Other; |
| SQ | |

Alignment Scores:
Pred. No.: 1.33e-09 Length: 103599
Score: 614.50 Matches: 667
Percent Similarity: 31.75% Conservative: 325
Best Local Similarity: 21.35% Mismatches: 1125
Query Match: 4.65% Indels: 1014
Gaps: 154

US-09-522-753-5 (1-2517) x ABX04971 (1-103599)

| | | | |
|----|-------|--|-------|
| Qy | 9 | AlaGlnThrTrpArgAlaThrGluProArgTyrProHisSerLeuSerTyrProVal | 28 |
| Db | 69260 | GCTACGCGCTTGTCTTAGC---CCACGCTCCGCCACACCCC---AGCAATCGGATC | 69207 |
| Qy | 29 | GlnIleAlaArg-----ThrHisThrAspValGlyLeuLeuGluTyrGlnHis--- | 44 |
| Db | 69206 | CACCTCACCGCAGCGACCTGAAATCACCTCGAGTCAGGTGTGTGACCGCCCACT | 69147 |
| Qy | 45 | HisSerArgAspTyrAlaSerHisLeu----- | 53 |
| Db | 69146 | CACGACGA-----GCCATCCACATCCGCGGCGAGCTCGGCGGTGCGCTGCACCGC | 69093 |
| Qy | 54 | -----SerProGlySerIleIleGlnProGlnArgArgArgPro | 66 |
| Db | 69092 | ACACATCGCCCCCTCAGGCAACCCACCCATCAACGCGCGCGCCGCCACCCCGC | 69033 |
| Qy | 67 | SerLeuLeuSerGluPheGln---ProGlyAsnGluArgSerGlnGluLeu--- | 82 |
| Db | 69032 | ACACGATCCGCCAGATCAAGACCCCGCCCATGCGCGCGCGCGGATCTCACCGATCGA | 68973 |
| Qy | 83 | -----HisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer | 98 |
| Db | 68972 | ATGCCCGGACGACCATCCCGCGCACCCC---GACCGACTCCACAAACCGGGCCACCC | 68916 |
| Qy | 99 | GluMetGluPheIleGluSerLysArgProArgLeuLeuLeuLeuProAspProLeuLeu | 118 |
| Db | 68915 | CAC-----CTGCAACGCAACAAACCCCGCTGCGC-----CCACCGGTGTGATC | 68871 |
| Qy | 119 | ArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp | 138 |
| Db | 68870 | CAACCGCTCCCG----- | 68859 |
| Qy | 139 | ArgSerLeuThrGlyLysLeuGluProValSerProProHisThrAspPro | 158 |
| Db | 68858 | CGGGCCCCCAAGAACACCCCTC-----CCGCNAACCCACCCGCAACCAACCCC | 68814 |

QY 611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHis 630
Db 396 -----ACTACAGAGCTCTCTACACAGCACCGGAGCAGCCATCACCC 440
QY 631 GlyArgAsnTrpSerAla-----IleAlaArgMetValGlySerLysThrValSer 647
Db 441 AGCACACCATGGCGGGCGTGAACCGGGCGCTCATCGCCAAACCGCTCTCGTACACCTCG 500
QY 648 GlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGln 667
Db 501 GCCTGCAGG----- 509
QY 668 GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaPro 687
Db 510 -----GCCCGAGCCTCACCGCTCGACCGCGCAGTCTCTCGCTCGTCCGCG 557
QY 688 AlaAlaAlaSerGluGluAlaAlaPheProValValGluAspGluGluMetGluAla 707
Db 558 TGCACCTGGCCTCGAGTCCCTCGCGCGCGGAGTCCACGACGGCGCTCGTCCGCGCG 617
QY 708 SerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSer 727
Db 618 TGAACCTCAACATCTCCGCGAGAGCGCTGAGCGGAGGAGCTTCGGTGGACTCTCCC 677
QY 728 GlyAsnGluValProArgGlyGlyCysSerGlyPro-----AlaThrVal 742
Db 678 CGGACGCGCACCGCTCACCTTCGACGCGCGGCCCAACGAGTTCGTCGCGGCGAGGGCG 737
QY 743 AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThr 762
Db 738 CGCGAGTCTGCTGACTCAAGCCCTCTCCCGCGCTCTCCGCGCGCGGACCGTGTCCACG 797
QY 763 GlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyPro 782
Db 798 CGTCACTCCGCGCAGCGCGTCAACACACGAGCCACCCCGGGTCTCACCTGCGCCA 857
QY 783 ProThrProPro-----ArgArgThrSerArgAlaProIleGluPro 796
Db 858 GCAGGCGCGCCAGGAGAGTCTGCGCGAGGCGTACCGAAGCGCGCTCGGACCGCT 917
QY 797 ThrProAlaSerGluAlaThrGly---AlaProThrProProAlaProProSerPro 815
Db 918 CCGCCCTCCAGTACGTGCAACTCCAGCGCACCGGAACCCCGCGCGG-----ACCCCA 971
QY 816 SerAlaProProValProValProLysGluLysGluGlu----- 830
Db 972 TCAGGCGCGCGCTCGCGCGCTCTCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1031
QY 831 ThrAlaAlaAlaProProValGluGluGluGluGluGluGluProProAlaAlaGluGlu 850
Db 1032 TCGTGGCTCGGCAAGACGACGTCGGGACCTCGGACACCTCGAAGCGCGCGCGCATGTCGCGC 1091
QY 851 LeuAlaValAspThrGlyLysAla----- 858
Db 1092 TCATCAAGACGCTCTCTCGCTCGCGCGCGCGCGATCCCGGAGCGCTCAACTCCGTA 1151
QY 859 -----GluGluProValLysSerGluCys----- 866
Db 1152 CGCCCGCACCGGACATCCCGCTCGACACCTCGGGCTCGAGTTCGCGAGCGCGCTCGGG 1211
QY 867 -----ThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
Db 1212 AGTGGCGCGCACCGGACCGGAATCTCTCGCGCGCGTCACTCTGTCGGCATGCGCGCA 1271
QY 884 GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlySerGlyArgAla 903
Db 1272 CCAACGCCCACTGCTCTCAGGGAAGCGCCCGCCAGGCGCGGAGCAGCCCGCATCG 1331
QY 904 ThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAla 923
Db 1332 ATGAGGAGACCCCGCTCGACAGCGGGCGCGCACTGCGCTCTGCTGCTACCGCGCGCGCG 1391

QY 924 AspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSer 943
Db 1392 -----GCAGGCGCTCTGCGCGCGCGCGCGCG 1418
QY 944 LeuLeuThr---ProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp 962
Db 1419 GCTGACAGGCGCGTCTGAAGCGGACCGGAGTCTCGCGCGCGCGCGCGCTCGCGCG 1472
QY 963 LeuLysGlnLeuLysGlnArgAlaAlaIleProProIleGlnValThrLysValHis 982
Db 1473 -----GTCGCTGCTCACCGCTAGCGTCTTACCGCACCGGT 1511
QY 983 GluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProPro 1002
Db 1512 CGGTCTCTCTCGCGCGCGCGCGCTCTCTCGACGCGCTCGCGCGCGCTCGCGCGCG 1571
QY 1003 GlnAsnLeuGlnProGlu---SerPheAlaProGlnGlnProGlySerSerProArgGly 1021
Db 1572 GGACGCGCGCGCGCGGTGTCTACCGGACCGCGCGCGCGCGCGCTCTCGCGCTCTGT 1631
QY 1022 LysSer-----ArgSerProAlaPro 1028
Db 1632 TCAGCGCGCGGTCGCCAACGTACGGGATGGGCATGGAGTTGTACGCGCGCGCGCG 1691
QY 1029 ProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProCys 1048
Db 1692 CTTTCGAGCGCGCTTCGACGCGCTCG-----CGCGCGAAC 1727
QY 1049 TrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerProHis 1068
Db 1728 TGACCGCTCTCTCGACCGCGCGCG-----TCGCGGAATCTCTCGCGCGCGCGCGCG 1784
QY 1069 ---AlaProAspProSerAlaPheSerTyrAlaPro-ProGlyHisProLeuPro----- 1085
Db 1785 ACCGACCGCTCACACACAGCGCGCTCTTCGCGTGGAGTTCGCGCTCCCGCGCTCG 1844
QY 1086 -----LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIle 1103
Db 1845 TCGAGTCTCTGGGCGGTGAC-----GCCGACCTCTCTCGCGCGCGCGCGCGAG 1897
QY 1103 rAsnProProProLeuLysSerSerAla-----LysHisProSerValLeuGluAr 1120
Db 1898 ATCAGCGCGCGCGCGCTCTCTCGCGGCTCTCTCTCGCGCGCGCGCGCGCGCTCTCTCGCG 1957
QY 1120 gGlnIleGlyAlaIleSerGln-----GlyMetSerValGlnLeuHisVa 1135
Db 1958 GCGCGC-GGCGCGCTCATCGAGCGCTCCCGAGGCGCGCGCGATGTCGCGTTCGAGGC 2016
QY 1135 lProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuPr 1155
Db 2017 GAGCGAGGAGGAAGT-GCTTTCGCGACCTCGCGGACGCGAGCGGAGCTCTCCCT----- 2070
QY 1155 oMetAspProLysLeuAlaProPheSerGlyValLysGlnGluLeuSerProAr 1175
Db 2071 -----CGCGCGCGTGAACCGCGCGCGCG 2093
QY 1175 gGly-----GlnAl 1178
Db 2094 CGGTCTCTCTCGCGCGCGCGCGCTCTCTCGACGTCGCGCGAGCTCTCGCGCGAAC 2153
QY 1178 aGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh 1198
Db 2154 AGGCGCGCGGAGGAGCGCTCAGCGTCTCGCACCGCTTCCACTCCCGCTCATGGAGC 2213
QY 1198 rAlaLeuGlySerValProGlySerIleThrLysGlyIleProSerThr----- 1215
Db 2214 CGATGCTCGACGACTTCCCGCGGTCGTGGAAGAGCTGAGCTTCCAGGAGCGCGCGCTCG 2273
QY 1216 ---ArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrPr 1234
Db 2274 ACCTCTGTCTCCAGGTGAC-----GGGCTGCTGTGTACAGCGCGCG 2315
QY 1234 oAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerAr 1254

Db 6015 -----GCGACCACTCGGGCGGTCCGGGTGGAGAACTCACCTCGAAGCGCCGC 6065

Qy 2336 AlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAsp 2355

Db 6066 TCGTCTCTCCGCGAGCGGGCGCGTCCGCGATCCAGGTCCGCGAGCGGCGAGT 6125

Qy 2356 GluTrpGluSerProLeuSerAlaAsnAlaPheAsnPro-----Leu 2371

Db 6126 CGCGCGCGGGCGACCTTCGGTGTGTACAGCACCCCGGCTCGGCGACACCGGTGACG 6185

Qy 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaPheGlyA-gSerAsp 2391

Db 6186 AGCGCGCGCGGAGTACCGCGCATGTCTCCGCGTACTCGCGAAGGGGACCGGCCA 6245

Qy 2392 HisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArg----- 2408

Db 6246 CGGAGTCGGACCAACCGCGGACCGACCGGAGCGGTTCAGCGGCTCGCGCGCTCGCGCGG 6305

Qy 2409 -----Pro-SerSerArgLysAlaLysSerProAlaProGly-----LeuAl 2422

Db 6306 CGACCGCACACCCCTCGACGGCGTCTACGACCGGCTCGCGGAGCTCGGCTACGGATACG 6365

Qy 2422 aSerGly-----AspArgPr 2427

Db 6366 GTCCGGCTTCCAGGCTGACGGGCTGTGGCGGCGGCGCACAGCTCGCCGAGA 6425

Qy 2427 oProSerValSerSerValHisSerGluGlyAspCysAsnArgThrProLeuThrAs 2447

Db 6426 TCCGGCTCGCGCGCGCACGACGAGCGCGGGCTTTCGGCGTACACCGCGCGTGC 6485

Qy 2447 nArg-----ValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAs 2465

Db 6486 TCGAGCGGGCGCTCACCGGATCTCTCTGGAGG-GCACTAGTGGCGGTCTGTGAC 6544

Qy 2465 nProLeuLeuMetArgLeuGluAlaGlyValMetAlaSerProProProGlyLeuPr 2485

Db 6545 GCGGATA-----CCGACGCGACC 6562

Qy 2485 oAlaGlySerGly-----ProLeuAlaGlyPr 2494

Db 6563 GACCGGATCCGGCTCCGCTGGCGGGGTGAGTCCCTCCACGCGGAGGGGCCACC 6622

Qy 2494 oHisHialaTrpAspGluGluPro 2502

Db 6623 GCGTCCGCTACGGATCACACCA 6647

RESULT 74

ID AA287283 standard; DNA; 15872 BP.

XX AA287283;

AC

XX 15-SEP-2003 (revised)

DT 05-JUN-2000 (first entry)

XX

XX S. venezuelae vep ORF 1, SEQ ID NO:1.

XX

XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;

KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;

KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;

KW chronic obstructive pulmonary disease; respiratory inflammation;

KW hypercholesterolaemia; crop protection agent; ds.

XX Streptomyces venezuelae; ATCC15439.

OS

XX Key Location/Qualifiers

FH 20. .13912

CDS

FT /*tag= a

FT /product= "vep ORF 1 amino acid sequence #1 (AAV77177)"

FT 14056. .14151

FT /*tag= b

FT /product= "vep ORF 1 amino acid sequence #3 (AAV77199)"

FT 14167. .15827

FT CDS

FT /*tag= c

FT /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"

XX

PN W0200000620-A2.

XX 06-JAN-2000.

PD

XX 25-JUN-1999; 99WO-US014398.

XX

XX 26-JUN-1998; 98US-00105537.

PR

XX (MINU) UNIV MINNESOTA.

XX Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI; 2000-160679/14.

DR P-PSDB; AAY77177, AAY77178, AAY77199.

XX

XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

PT synthesis of methymycin and pikromycin.

PS Example 3; Fig 23; 438pp; English.

XX

The invention relates to an isolated and purified nucleic acid segment comprising a desosamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the eryC gene cluster of *Saccharopolyspora erythraea* or *Streptomyces antibioticus*. The invention also relates to a macrolide biosynthetic gene cluster, or fragments thereof. The macrolide biosynthetic gene cluster encodes proteins which synthesise methymycin, pikromycin, neomethymycin, narbomycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthase may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. The present sequence represents a *Streptomyces venezuelae* ATCC 15439 DNA sequence, designated vepORF 1 in the specification, which actually contains 3 open reading frames, which encode proteins AAY77177-Y77178 and AAY77199. The vep ORF 1 protein is defined in the specification as a PHA monomer synthase. (Updated on 15-SEP-2003 to standardise OS field)

XX SQ Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.15e-10 Length: 15872

Score: 614.50 Matches: 520

Percent Similarity: 31.11% Conservative: 214

Best Local Similarity: 22.04% Mismatches: 964

Query Match: 4.65% Indels: 670

DB: 3 Gaps: 101

US-09-522-753-5 (1-2517) x AA287283 (1-15872)

Qy 571 ThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAla 590

Db 297 TCTTCGAACCTCGCTGGGAGCGCG---TGGAGGACGCCGGAATCG-----TCCCGCGCA 347

Qy 591 AsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsn 610

Db 348 CCCTCGCGGAGCCGACCGCGCTCTTCGTCGCGCACCTCGCGGACG----- 395

| | | | | |
|----|------|--|-----------------------|------|
| Db | 2071 | ----- | CGGGCGCGTGAAGCGCCCGCG | 2093 |
| Qy | 1175 | gGly- | -----GlnAl | 1178 |
| Db | 2094 | CGTCTCTCGGGCGCGCGCGCGTCTCGACGTCGCGAGTCTGCGCGAAC | 2153 | |
| Qy | 1178 | aglyProPrgluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh | 1198 | |
| Db | 2154 | AGGCGCGCGGACGAAGCGCTCAGCGTCTCGCACGCTTCCACTCGCGCTCATGGAGC | 2213 | |
| Qy | 1198 | rAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThr | 1215 | |
| Db | 2214 | CGATGCTCCAGCACTTCGCGCGGTCTCGAAGAGCTGGACTTCCAGGAGCCCGCGTGC | 2273 | |
| Qy | 1216 | ----ArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrPr | 1234 | |
| Db | 2274 | ACGTCGTGTCCACGGTGAC | 2315 | |
| Qy | 1234 | oAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerAr | 1254 | |
| Db | 2316 | AATGAC | 2357 | |
| Qy | 1254 | gLeuAspArgGlyArgGluAspSerLeuProLysGly | 1270 | |
| Db | 2358 | CCGTACGCTTCTCGACCGCTAGCCCTCGAGGATCGGGCGCGACACCTTCTGCG | 2417 | |
| Qy | 1270 | rGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe | 1290 | |
| Db | 2418 | AGCTCGGTCCCGACGGGTCTGCTCCGCGATGGCGGGACTCCGTACGCGACGAGG | 2477 | |
| Qy | 1290 | rLysGluAspGly | 1309 | |
| Db | 2478 | CGGCACCGCGGTCTCGCCCTCGCAAGGCCCGCGAGCCCGACCGTCTGCTGCCCG | 2537 | |
| Qy | 1309 | gThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluI | 1329 | |
| Db | 2538 | CACACACACGTCTT | 2562 | |
| Qy | 1329 | yLeuMetGlyArgAlaIleProProGluArgHisSerProHisIleLysGlyGlnHi | 1349 | |
| Db | 2563 | -----CCACGACGTCTGACCGCGCGCGCGCGCGGAGCACCG | 2600 | |
| Qy | 1349 | sHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAs | 1369 | |
| Db | 2601 | GCACGTCAGGTCGCCC-TGCCGHACTACGCTTCCAGCGCGAGCCACTGTTTCGAC | 2659 | |
| Qy | 1369 | pTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProPr | 1389 | |
| Db | 2660 | GGCGCGCGGAAACGGCGCCCGCTCACGGCGGCCGATCGGGCACCGGTGCGGGCAC | 2719 | |
| Qy | 1389 | oSerArgAspLeuThr | 1408 | |
| Db | 2720 | GGCGCGCGCGGTGTGACGTCGGCGGCGGCGGAGGGCGGCGCGCGGTGCG | 2779 | |
| Qy | 1408 | sProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluI | 1428 | |
| Db | 2780 | GTTGGCGGTGATCGCGCGCTCGCCACG---AGACGACCGAGCGGTGCGCGCACACGTC | 2836 | |
| Qy | 1428 | eProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysG | 1448 | |
| Db | 2837 | GGCGCGCTCTCGAGTACACGA---CCGAC---CCGCGTCGAACCTCGGCTCACCTTC | 2890 | |
| Qy | 1448 | luGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlys | 1468 | |
| Db | 2891 | AAGGAGCT | 2941 | |
| Qy | 1468 | erLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValH | 1488 | |
| Db | 2942 | GTCAGACAC | 2983 | |
| Qy | 1488 | isProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerL | 1508 | |
| Db | 2984 | GACCACCGACCGCGCGCCCTCTCGCCGCCACCTGGGGCGACCTGCTCACCGCGGCGACG | 3043 | |
| Qy | 1508 | eLysSerArgProGly | 1524 | |
| Db | 3044 | GGCGAGACCGGATCGCCGACCGGATACCGCGCGGACCCCGCGGACACACCGCGCGAG | 3103 | |
| Qy | 1525 | -----GlyAlaProValIleValProGluLeuGlyLysProArgG | 1538 | |
| Db | 3104 | CCCATCGCATCATCGGCATGCGCTGCG---CTACCCCGCGCGGTCACTCCCTCC | 3157 | |
| Qy | 1538 | InSerProLeuThrTyrGluAspHisGlyAlaPro | 1555 | |
| Db | 3158 | GAGGACCTGTGGCGCTCTGCGCAGAGGGCGGACGCGCTCTCGGGGCTGCCACCGAC | 3217 | |
| Qy | 1555 | roArg | 1571 | |
| Db | 3218 | CGCGGTGGAGCAGGACCTCTTCAGACCGGACCCCGACGCGCAGCGCAGAGCTCGGTC | 3277 | |
| Qy | 1572 | ----GlySerLeuSerSerLysAlaSerGlnAspArgLysLeu | 1585 | |
| Db | 3278 | CGCGAGGGCGGATCTCTGACGACGCGCCCTGTTTCGACGCGCGCTTCTTCGGGATATCG | 3337 | |
| Qy | 1585 | ----- | 1585 | |
| Db | 3338 | CCCGCGAGGCCCTCGGCATGACCCGACGAGCGGCTGCTCTCGAGACGCGCATGGAG | 3397 | |
| Qy | 1586 | -----ThrSerThrProArgGluIleAlaLysSerProHisSerThrValP | 1601 | |
| Db | 3398 | GCGGTGGAGCGCGAGGGCTCGACCCCGAAGCCCTCAAGGGCAGCGGACCGCTTTC | 3457 | |
| Qy | 1601 | roGluHisHisProHisProIleSerPro | 1620 | |
| Db | 3458 | GTCCGCGCACCGCCCTGACACTACGCGCGCGCATACACGCGCGCGCGGCGCTCGAG | 3517 | |
| Qy | 1620 | yValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro | 1638 | |
| Db | 3518 | GGCCACTCTGACCGGGACCGCCCGGTGATGTCGGCGCGCATCGCTACACAGCTC | 3577 | |
| Qy | 1639 | -----ArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuA | 1656 | |
| Db | 3578 | GGCTCACCGGTCTCGGTCTACCGTCGACACGCGCTGCTCTGCTCTGCTCGCGGTG | 3637 | |
| Qy | 1656 | laProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspT | 1676 | |
| Db | 3638 | CACCT | 3655 | |
| Qy | 1676 | hrAlaAla | 1693 | |
| Db | 3656 | CTCGGCGAGCGCGAGTCTGAGCTCGCGCTCGCGCGGAGCGACCGTCTATCGACACCG | 3715 | |
| Qy | 1693 | lnGlnMetHisHisAsnThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyL | 1713 | |
| Db | 3716 | GGCATGTTCTCGAGTTCGCGGCGAGCGGCTCGCGCGCGACCGCGCGCTCCAGGCC | 3775 | |
| Qy | 1713 | eu | 1718 | |
| Db | 3776 | TCTCCGACTCCGCGCAGGCACCTCTGCGCGCGAGGGCTCGGCTCTCTCGTCTCGAG | 3835 | |
| Qy | 1718 | erSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValP | 1738 | |
| Db | 3836 | CGGCTCTCGAGCGCGAGCGCAACCGGCACCGCTGTCGCCGTGATTCGCGGCGAGCGCG | 3895 | |
| Qy | 1738 | roHisLeuProValLeuValProProThrProGlyThrProAlaThrAla | 1754 | |
| Db | 3896 | GTCAA | 3951 | |
| Qy | 1755 | -----MetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer | 1770 | |
| Db | 3952 | CGCGTCTATCCGACAGCCCTTGGCGCGAGCGGCTCACCCCGCGCGACGCTCGACGCGCT | 4011 | |
| Qy | 1770 | erArgHis | 1783 | |
| Db | 4012 | CGAGCGCGACCGGTACCGGCTCGGCGACCCCATCGAGGCCCGCGATCTCTCGG | 4071 | |

SQ Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 U; 0 Other;

Alignment Scores: 2,94e-10 Length: 15872

Pred. No.: 615.50 Matches: 524

Score: 31.11% Conservative: 208

Best Local Similarity: 22.27% Mismatches: 972

Query Match: 4.66% Indels: 658

DB: 2 Gaps: 102

US-09-522-753-5 (1-2517) x AAT68715 (1-15872)

```
QY 571 ThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAla 590
DB 297 TCCTGAACCTCGCTGGAGGCGC---TGGAGACCGCGAATCG-----TCCCGGCA 347
QY 591 AsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsn 610
DB 348 CCTCTCCCGAAGCCGACCGCGCTTTCGTGGCCACCTCGCGGAGC-----395
QY 611 GluSerSerArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHis 630
DB 396 -----ACTACAGAGCTCTCTTACCAGCAGCGGAGCCATCACCC 440
QY 631 GlyArgAsnTrpSerAlaIleAlaArgMet-----ValGlySerLysThrValSer 647
DB 441 AGCACACCATGGGGCGGTGAACCCGGGCTCATCGCAACCGGCTCGTACACCTCG 500
QY 648 GlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGln 667
DB 501 GCCTGCAGG-----509
QY 668 GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaPro 687
DB 510 -----GCCGAGCTCACCCTGCAGCCGCGAGTCTCTCGCTCGTGGCG 557
QY 688 AlaAlaSerGluGluAlaAlaPheProValValGluAspGluMetGluAla 707
DB 558 TGCACCTGGCTCGAGTCTCTCGCGCGCGGGAGTCCACGAGCGGCTCGTCCGCGCG 617
QY 708 SerGlyValSerGlyAsnGluGluMetValGluGluAlaLeuHisAlaSer 727
DB 618 TGAACCTCAACATCTCGCGGAGAGCGCGTGAAGAGGAGCGTTCGGTGAAGTCTCC 677
QY 728 GlyAsnGluValProArgGlyGluCysSerGlyPro-----AlaThrVal 742
DB 678 CGGACCGCACCGCTACACTTCAGCGCGCGGCCAACGATTCTGTCGGGGCGAGGCG 737
QY 743 AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaLysAspThr 762
DB 738 GCGGAGTCTGCTACTCAAGCCGCTCTCGCGCCCTCGCGAGCGGACCGGTGTCCAG 797
QY 763 GlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyPro 782
DB 798 GCGTATCTCGCGCCAGCGCGTCAACACAGCAGGACCCCGGGTGTCAACGTCGCCA 857
QY 783 ProThrProPro-----ArgArgThrSerArgAlaProIleGluPro 796
DB 858 GCAGGCGCGCCAGGAGAGGTCTCGCGAGCGGTACCGAAGGCGGCGCTCGAGCCGT 917
QY 797 ThrProAlaSerGluAlaThrGly---AlaProThrProProAlaProProSerPro 815
DB 918 CCGCGCTCCAGTACGTGCAACTCCAGCGCACCGGAACCCCGTCCGCGC-----ACCCCA 971
QY 816 SerAlaProProValValProLysGluGluLysGluGlu-----830
DB 972 TCAGGCGCGCGCGCTCGCGCGCTCTCGGTCTCGGCGCGCCCGCGGACGAAACCCCTGC 1031
QY 831 ThrAlaAlaAlaProProValGluGluGluGluGluGluGluProProAlaAlaGlu 850
DB 1032 TCGTGGCTCGGCGAAGACGTCGGGACCTCGAAGCGCGCGGCGATCGTGGCC 1091
QY 851 LeuAlaValAspThrGlyLysAla-----858
```

```
DB 1092 TCATCAAGACGCTCTCGGCTCGGCCCGCGCGGATCCCGGAGCCTCAACTTCGTA 1151
QY 859 -----GluGluProValLysSerGluCys-----866
DB 1152 CGCCCGACCGGACATCCGCTCGACACCTCGGGCTCAGCTGCCGAGCGCTCGGGG 1211
QY 867 -----ThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAla 893
DB 1212 AGTGGCCGACCCGAGCGGAACTCTCGCGCGGTAGCTGTTCCGCGATGGCGGCA 1271
QY 884 GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGlyGlySerGlyArgAla 903
DB 1272 CCAACGCCACGTCGCTCTCAGCGAAGGCCCGCCAGGCGCGGACGCGGATCG 1331
QY 904 ThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAla 923
DB 1332 ATGAGGAGACCCCGCTCGACAGGGGCGGCGACTGCCCTTCGTCTGTCACGCGCGGCG 1391
QY 924 AspGluValAspGluAlaGluGlyAspLysAsnArgLeuLeuSerProArgProSer 943
DB 1392 -----GCGAGGCCCTCGCGCGCCAGCGCGGC 1418
QY 944 LeuLeuThr---ProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp 962
DB 1419 GCTGACGAGCGCTCGAAGCGGACCCGAGGCTCGCGCCGCGCACCTCGCCC-----1472
QY 963 LeuLysGlnLeuLysGlnArgAlaAlaIleProIleGlnValThrLysValHis 982
DB 1473 -----GGTCTGCTGTCACACCGTACGCTTCACGACCGGT 1511
QY 983 GluProProArgLysAlaAlaProThrLysProAlaProProAlaProProProPro 1002
DB 1512 CGTCTGCTCTCGCCCGGACCGCGCGCTCTCTCGACGCGCTCGCGCGCTCGCGCG 1571
QY 1003 GlnAsnLeuGlnProGlu---SerAspAlaProGlnGlnProGlySerSerProArgGly 1021
DB 1572 GGACGCGCGCGCGCGCGTGTGTCACCGGACCGCGCGCGCGCGCGCTCGCGCTCTGT 1631
QY 1022 LysSer-----ArgSerProAlaPro 1028
DB 1632 TCAGCGCGCAGGTGCCAACGTACGGGATGGGCATGAGTTGTACGCGGCCACCCCG 1691
QY 1029 ProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCys 1048
DB 1692 CCTTCGCGAGCGCTTCGACGCGCTG-----CGCGCAAC 1727
QY 1049 TrpThrSerGlyLeuProPheProValProProArgGluValLysLysAlaSerProHis 1068
DB 1728 TGGACCCCTCTCTCGACCGCGCCC---TCGCGCAACTCTCGCGCGGCGGACACCTCG 1784
QY 1069 ---AlaProAspProSerAlaPheSerTyrAlaPro-ProGlyHisProLeuPro-----1085
DB 1785 ACGGACCTTCACACACAGCGCGGCTCTTCGCGGTGAGGTGCGCTCCACCGCTCG 1844
QY 1086 -----LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSe 1103
DB 1845 TCGAGTCTTGGGCGTTCAC-----GCCGACCTGTCTCG-----CGGCGCACCTCGTGGCGAG 1897
QY 1103 rAsnProProProLeuLysSerAla-----LysHisProSerValLeuGluAr 1120
DB 1898 ATCAGCGCGCGCCACGTCGCGGGGTCTCTCGCTCGCGACGCGCGCGCTCGTGGCG 1957
QY 1120 gGlnIleGlyAlaIleSerGln-----GlyMetSerValGlnLeuHisVa 1135
DB 1958 GCGCGC-GGCGCTCTATCAGCGCTCCCGGCGCGCGGATGTCGCGTTCGAGGC 2016
QY 1135 lProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuPr 1155
DB 2017 GAGCGAGGAGGAAGT-GCTTTCGCACTTCGCGGACGCGAGCGGAGCTCTCCCT-----2070
QY 1155 oMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSerProAr 1175
```

QY 2137 nHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAl 2157
 Db |||||
 5008 -----GACAGGAAACCAACCGCTCAACAGC 4983
 QY 2157 aPro---LeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLe 2176
 Db |||||
 4982 GCCGGAATCGCGTGGATCGTCCGTTCCCGACGAGCGCGCTGTCCAAATCAGCGGCA 4923
 QY 2176 uArgArgProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaArgG1 2196
 Db |||||
 4922 ACAGCGCAACGCC-----GGTCTTCGCGA-- 4898
 QY 2196 ySerProHisSerGluGlyGlyLysArgSerProGlu-----ProAsnly 2211
 Db |||||
 4897 -ACGCGCATCGAACAGCGGAGTCCATCTCCGTGGMAAGTTCTGCGATGCCGCTCC-- 4841
 QY 2211 sThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMe 2231
 Db |||||
 4840 -----GCGCCCAACCGCTCCACCG-----AC 4821
 QY 2231 tThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluG1 2251
 Db |||||
 4820 GCCCGCTCCAAAGCCACTCGTCA-----TACCGCTGGGTTCGGCC 4782
 QY 2251 nThrGluPro-----SerArgMetGlySerLysSerProGlyAsn----- 2264
 Db |||||
 4781 CACAGTCCCCACGCCAACACACACCCAGGACGCCCGCGCGCGATCAACGCGCAAT 4722
 QY 2265 -----ThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAl 2281
 Db |||||
 4721 GCATCCAGGAACGTGTTCCGCGCGCGGTAGTGTGCGCTGCGGACCGCAATCCCG 4662
 QY 2281 aMetValLys--SerLysLysGlnGlu----- 2289
 Db |||||
 4661 GAGAGGAGAGAGAAAGAACGACAGACAGATCCAGATCGCGGTCAACTCGTCGAGA 4602
 QY 2290 -----IleAsnLysLysLeuAsnThrHis-----AsnArgAsnGluP 2302
 Db |||||
 4601 TGCCATGCCCGCGTCACCTCGGGGCAACACCCGTGCCCATCCGCGCGCGCAAGGAC 4542
 QY 2302 roGluTyrAsnIleSerProGlyThrGluIlePheAsnMetPro-AlaIleThrGly 2321
 Db |||||
 4541 CCGATCACTCGTCGCGGACGACCCAGCGGTATGACAGATCCCGCGCGCGCGCGAGTCG 4482
 QY 2322 ThrGlyLeu-----MetThrTyrArgSerGlnAlaValGlnGluHis 2335
 Db |||||
 4481 ACCGGAATGGCGGCCAGCAGCTGCTCCAGATCGGTGCGGTCCAGCCACATCCGAGCCACC 4422
 QY 2335 AlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAsp 2355
 Db |||||
 4421 ACCTCAACCAACG----- 4410
 QY 2356 GlnTyrGluLysSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
 Db |||||
 4409 -----GCGCGCGCGCGCGAGCTCATCCACCACTCGTGAGCTCCAGCGCGCA 4362
 QY 2376 SerLeuPro-AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuTh 2395
 Db |||||
 4361 TTCCAGCCACGCGCGCTGCCAACCAACCGCAATCCCGTCT----- 4316
 QY 2395 rSerProGlyGlyGlyLys----- 2402
 Db |||||
 4315 -CTCCACCAAGGTGACGGGCAACCACTCCCGCCACGACCGGTGCCCGCGGTGATGAGG 4257
 QY 2403 -----AlaLysValSerGlyArg----- 2408
 Db |||||
 4256 ACGGTCCGCTCCGTCGAGTTGCGGGGAGGAGCCCTCTCCGCGACCGGTGAGTCGCGCC 4197
 QY 2409 -----ProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAs 2425
 Db |||||
 4196 AACCGAGGACCAACGCGCGCGCGGAGTTCCTGCTCGCAGC 4137

QY 2425 pArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLe 2445
 Db |||||
 4136 CGCGCG-----GCAACGCGCTGC 4119
 QY 2445 uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAs 2465
 Db |||||
 4118 CACGACTCAGGTGTGCGGTCCACGTCCACGACGACGACGACGACGACGACGACGACGAC 4068
 QY 2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuPr 2485
 Db |||||
 4067 TCCGACTGGGTGAACTCAACACCCCGACGACGACGACGACGACGACGACGACGACGAC 4008
 QY 2485 oAlaGlySerGlyProLeuAlaGlyProHisHisAla 2497
 Db |||||
 4007 GAATCCGAGTCGCGCGACA-----CCGCTGCACGCG 3977

RESULT 73
 AAT68715
 ID AAT68715 standard; DNA; 15872 BP.
 XX
 AC AAT68715;
 XX
 DT 01-SEP-1997 (first entry)
 XX
 DE Streptomyces venezuelae polyketide synthase vep ORF1.
 XX
 KW Polyketide synthase; polyhydroxyalkanoate monomer synthase;
 KW polyhydroxybutyrate; biodegradable polymer; vep gene;
 KW metabolic engineering; ss.
 XX
 OS Streptomyces venezuelae.
 XX
 FH Key
 FT CDS 20..13912
 FT /tag= a
 FT CDS 14036..14136
 FT /tag= b
 FT CDS 14148..15827
 FT /tag= c
 XX
 PN W09722711-A1.
 XX
 PD 26-JUN-1997.
 XX
 PF 18-DEC-1996; 96WO-US020119.
 XX
 PF 19-DEC-1995; 95US-0008847P.
 XX
 PR (MINU) UNIV MINNESOTA.
 XX
 PA Sherman DH, Williams MD, Xue Y;
 XX
 PI WPI: 1997-341701/31.
 XX
 DR P-PSDB; AAW19629, AAW19630, AAW00918.
 XX
 DR Expression cassettes for production of poly(hydroxyalkanoate(s) -
 XX provide wide range of biodegradable polymers for medical or industrial
 XX use.
 XX
 PS Claim 54; Fig 23; 91pp; English.
 XX
 CC Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
 CC synthase (PKS) gene cluster encoding a polyene of 12 carbons (see also
 CC AAW19629-30 and AAW00918). It contains 5 PKS modules, with a 5' loading
 CC module and a 3' end domain. Each of the sequenced modules includes a keto
 CC -ACP, an acyltransferase, a dehydratase, a keto- reductase and an acyl
 CC carrier protein domain. The gene cluster was cloned using a heterologous
 CC hybridisation strategy from a genomic DNA library. A novel expression
 CC cassette encoding the first module from the vep gene cluster and module 7
 CC from the Streptomyces ty1P gene cluster has polyhydroxyalkanoate (PHA)
 CC monomer synthase activity and can be used for PHA prodn. in host (esp.
 CC insect) cells for use as a biodegradable polymer

| | | | |
|----|------|---|------|
| Qy | 1317 | gValGlyA-gAla-lleSerSerAla-SerlleGluGlyLeuMetGlyArg----- | 1333 |
| Db | 8010 | GGATGTGAGTCAAGCCAGTACCCTCGTCGTTGAAGGCATACGTGGCAGCTGCACCTG | 7951 |
| Qy | 1334 | -----AlalleProPro----- | 1337 |
| Db | 7950 | CTTTGCCCGGTACCGCGAAACAGCCGCAAGTCCACTCAGCACACGCACATGGAT | 7891 |
| Qy | 1338 | -----GluArghHis | 1340 |
| Db | 7890 | CTGGCCAATGCCGTGTTCACCGTGTGCGCTCGTCGCGATTGTCGCATGAGCGGAAC | 7831 |
| Qy | 1341 | SerProHis-----HisLeuylsegluin | 1348 |
| Db | 7830 | CGGGCCACCCCGTCCCGCTTGATCGGATCCGCGCGCATCTCTCGACACGCGCGACA | 7771 |
| Qy | 1349 | HiHisIleArgGlySerlleThrGlnGlylleProArgSerTyVaigluAlagInglu | 1368 |
| Db | 7770 | CACCCCATCCGACCACTCGACGAGGTGC-TCACACCGTAGCCGCGAGCGCAGCAA | 7712 |
| Qy | 1369 | AspTyLeuArgGluAlaLysLeuLysArgGluGlyThrProProProPro----- | 1387 |
| Db | 7711 | --CACCGTCGGGAAGCAGCGGGTCTC--GCACCTCGCGCCACCAATACTCCGCG | 7658 |
| Qy | 1387 | ----- | 1387 |
| Db | 7657 | TAGCCCGATGCCGACCTCATCGAGTCAACCGTCAACGTCGACAGATCGCAGCCCTG | 7598 |
| Qy | 1388 | -----ProProSerArghAspLeuThr | 1394 |
| Db | 7597 | ACGACCGGTAGTCCACGCTCCGCGCGATTTCGGGTGAATCCGCGCAGCATCGGATCCAT | 7538 |
| Qy | 1395 | GluAlaTyrlsThrGlnAlaLeuGlyProLeuLysLeuLysProAlahisGluGlyLeu | 1414 |
| Db | 7537 | GATCGAATGAACGCATGCG-----AAACCCGCAACCACTCGAAC | 7496 |
| Qy | 1415 | ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg | 1434 |
| Db | 7495 | GGATCCCTTGGCCGTCAGCGCGACCGGCAATGTCTACGACACATCCCGATCAGCGAGA | 7436 |
| Qy | 1435 | HisThrProGluLeuProLeuAlaProArg----- | 1444 |
| Db | 7435 | GCACCAACCGCCCCGGGAGCGTTGACCGCGGATGCCACACATCAGCGAGTCGGCCA | 7376 |
| Qy | 1445 | -----ProLeuLysGluGlySerIleThrGlnGlyThrProLeuLys | 1458 |
| Db | 7375 | GCAACGGCGCAGCTGCTCTCTCCAGCAGCGCAGCAGCATGCGACCGCCAGTTG--- | 7319 |
| Qy | 1459 | TyrAspThrGlyAlaSerThrThrGly-----SerlysHisAspVal | 1473 |
| Db | 7318 | ----GAGGGCTTGATCAACCGGGCAGCGCCCGCCACACCGCAGCGCATCCGCA | 7265 |
| Qy | 1474 | ArgSerLeulleGlySerProGlyArgThrPheProProValHisProLeuAspValMet | 1493 |
| Db | 7264 | ACGACAACACTC-----CAGCCCGGAACCGCGCCGCGAGCTCAC-----CGA | 7223 |
| Qy | 1494 | AlaAspAlaA-a-gAlaLeuGluAraAlaCystTyrlGluSer----- | 1507 |
| Db | 7222 | CCGAGTGGCCAGCACCAACCGGGCGGACACCCACAGAACCCCAACATCCAGAGTC | 7163 |
| Qy | 1508 | ---LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerlleAlaArgGlyAla | 1526 |
| Db | 7162 | CGACTTGMACCGGAACGAGGCCGCTCGGCCCAACAGGCTCGT----- | 7118 |
| Qy | 1527 | ProValileValProGluLeuGlyLysProArgGlnSerProLeuThrTyrlGluAspHis | 1546 |
| Db | 7117 | -----CCAGAAGTTGCGTATCCGAACCGGACACACATCTCGTAGCGGGGCCA | 7070 |
| Qy | 1547 | GlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThr | 1566 |
| Db | 7069 | TCTGCCCGAGATGCG---CATCCAGCTCGAGCAGGCTTCGTCAACCGCTCGCGGAACA | 7013 |
| Qy | 1567 | ProArgLeuGlnGluGlySerLeuSerSerSerIysAlaSerGln-----AspArgLys | 1584 |

QY 2438 pCyAsnAtqArqThrProLeuThrAsnArgValTrp-GluAspArgProSerSerAlaG 2458
 Db 1131 GTCATCGAGAGCGACCCATCCCGAGCGCGTGTAGTGACAAACATCTGTCGAGATTATCCAG 1072
 QY 2458 lySerThrProPheProTyAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaS 2478
 Db 1071 GA-----GCAC 1066

QY 2478 erProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaT 2498
 Db 1065 GGGATGAACCGCCAGCGGTGACCGAGCCCATGTCTCC-----CATGTTT 1021

QY 2498 rPaspGluGluProPro 2504
 Db 1020 GGGGCAAAACACCCAGCCCT 1001

RESULT 72

AAF88338/c

ID AAF88338 standard; DNA; 14775 BP.

XX AAF88338;

DT 28-AUG-2001 (first entry)

DE S. spinosa DNA fragment encoding ORF21, SEQ ID 47.

XX Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
 KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
 KW macrolide; insecticidal; polyketide synthase; ds.
 XX Saccharopolyspora spinosa.

XX DE19957268-A1.

XX 08-MAR-2001.

XX 29-NOV-1999; 99DE-01057268.

XX 27-AUG-1999; 99DE-01040596.

XX (FARB) BAYER AG.

XX Eberz G, Moehrlie V, Froese R, Velten R, Salas JA;

XX WFI; 2001-267102/28.

XX P-PSDB; AAB70968.

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 recombinant production of insecticidal spinosyns and their derivatives.
 PS Claim 7a; Page 239-264; 354pp; German.

XX This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
 CC identify, inactivate or modulate genes involved in the biosynthesis of
 CC (II); (ii) to generate a library of polyketide synthases; (iii) for
 CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
 CC aglycone; and (iv) for recombinant production of the corresponding
 CC enzymes, which are used for production of (II), their precursors or
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC macrolides with insecticidal, but not antibacterial, activity, and can
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence, ORF 21, encodes an S. spinosa polyketide synthase.

SQ Sequence 14775 BP; 1969 A; 4181 C; 5731 G; 2894 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.68e-10 Length: 14775

Score: 616.00 Matches: 636
 Percent Similarity: 32.03% Conservative: 313
 Best Local Similarity: 21.46% Mismatches: 1051
 Query Match: 4.66% Indels: 973
 DB: 4 Gaps: 136

US-09-522-753-5 (1-2517) x AAF88338 (1-14775)

QY 45 HisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArg 64
 Db 11463 CAGCAGCGCGCGTTCCTCCAGCACAGAGAGTGTCGCGAGTGAATACCCAAATCCAG 11404

QY 65 ArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeu 84
 Db 11403 CGGAGAGCGCCAGGACCCCGGTCCAGCGCGCTCGAGCGCAGCGC----- 11356

QY 85 ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGlu 104
 Db 11355 -----TTGGCGCTCGAGCGCTCGAGACTCTCTAGC 11326

QY 105 SerLysArgProArgLeuGluLeuLeu-----ProAspProLeuLeuArg 119
 Db 11325 CGAACATCCACGGGA-TCACCGCTGGATGCCAGTGCACCCCGCATCCGATGTGTGCGA 11267

QY 120 ProSerProLeuLeuAlaThrGlyGlnProAlaGly----- 131
 Db 11266 ---TCTCCGTGTTCGAGCGGTGTTCGAGGATGAGTGTGCTGTCGCGCTGA 11210

QY 132 -----SerGluAspLeuThrLysAsp---ArgSerLeuThrGlyLysLeuGlu 146
 Db 11209 TCCGGAACGACGACACCCCGCTCGAGAGGAGCGCGCTCTCGGCGAGGGGCTGTTC 11150

QY 147 ProValSerProProSerProHisThrAspProGluLeuGluLeuValProProArg 166
 Db 11149 CGGTGAGGAGCGAAGCGGCGCGGACACCATGTCCACCTGTGAGTGGGTGTTCACGT 11090

QY 167 LeuSer-LysGluGluLeu-----IleGlnAsnMetAsp-----Ar 178
 Db 11089 GCAACGTGCGAGGCAACTCCCGTGCAGCTGCGCTGCACCATCTTGTATGACCCCGCCA 11030

QY 178 gValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnG 198
 Db 11029 CACCGCGCGCGCTGCTGTCGCGCATGTGTGACTGTGATCGA----- 10987

QY 198 nGlnLeuGluGluAlaLysProGluProGluLysProValSerProProPr 218
 Db 10986 -----CCCCAGCACAGGCGCGCTCAGGATCCCG-- 10957

QY 218 olleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 238
 Db 10956 -----CTCCTGCCCATATGTTCGCGATCAGCGCTGCGC----- 10924

QY 238 aGluAlaAlaHisArgIleLeuGlu---GlyLeuGly----- 249
 Db 10923 -----CTCGATCGGATCCCGAGCGCGTACCGTCCGTCGCGCTCCACGACATCCA 10871

QY 250 -----ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrAr 264
 Db 10870 CATCGAGCGCGACAGCCCGCATTCGAGGGCGCTGCCGATCACCCTGCTGCGACG 10811

QY 264 gGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuLeuTyrPh 284
 Db 10810 G-----ACCATTCGCGCGCAGTCAGGCGCATTCGACGACCATCTCTGATTCACCG 10763

QY 284 eLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspG 304
 Db 10762 CAGACCGCGGAACACCGC-----CAACACCGCGGTGAC 10730

QY 304 nLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsnProArgArgArgAl 324
 Db 10729 CATTCGACGAGCATCCGACACACCGCTCCAG-----CAACACCGCGCGCACCTCAC 10676

QY 324 aLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluLeuArgLysGln-- 343

Db 3128 GCTTCGCCCGCG-----TGGAGCGCTCTCTTGGCCCGCTGACCCCGCGGGCGG 3081
QY 1872 GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGlu 1891
Db 3080 AGCGCTGCGCTCTCGCGCCAGGACACGTGCGCTACAGGTGAGCACCAGCGCGGC 3021
QY 1892 ProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgPro----- 1909
Db 3020 CCGGACCCCGGTCCCGCTCGCGCCGCGCTACCGGAGCGCGTCTCCACACCGTGG 2961
QY 1910 -----AlaAlaThrPhe-ProProAla----- 1916
Db 2960 ACGCTGCAAGACATGGCCCGCAGCGCTCCGCGCTCGGGCTCGGGACCCGACTTGG 2901
QY 1917 -----ThrHisCysProLeuGlyThrLeuAspGlyValThrProThrLe 1932
Db 2900 ACGCGGCGCGCTTCGGGACACCGCGCCCAACCGCTGGGGCTCGGGCCCGCTGC 2841
QY 1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
Db 2840 GGCCCGCTTCGTCTCTCGCGCGCGCGCGCTCGC-----CGAGCTCCGCGCGCCCG 2787
QY 1952 gAlaAspThrGlyHisAla-----PheLeuAlaLysProProAla-----ArgSe 1967
Db 2786 AGGCGCTCGCGCGAGCTCGCGCTCTTCGCGCGCGCGCTCTGGAGCCCGACCGCG 2727
QY 1967 rGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProPr 1987
Db 2726 AGCGCGCGCGGTGCTGGCGCGCGCGCGCC-TCGCGCGCGCGCGCGCGCGCGCG 2668
QY 1987 oValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAl 2007
Db 2667 GTGCGCTGGGACTTTCGCGCTTCGAGA-----CCAGCGTCC-----CGCGCGCGCGCG 2617
QY 2007 aSerProAspProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGl 2027
Db 2616 GCGCGCTGGAGACCCACCGG-CCCTCGCGCGCGCGCGCGCGCGCGCGGAGGA 2558
QY 2027 nSerLysProPheSerIleGlnGluLeuArgSerLeuGlyTyrHisGlySerSe 2047
Db 2557 CTCTCGCGCGGTGGAGATC-----GTGCGCTTCGCGCGCGCGGA 2519
QY 2047 rTyrSerProGluGlyValGluProValSerProValSerProSerLeuThrHisAs 2067
Db 2518 CGCGCGCGCGCGCGG-----CGCTCGCGCGCCATCAAGGTGGAGCCATCTCGGACGA 2465
QY 2067 pLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuAr 2087
Db 2464 CGA-GGAGCGGAGGAGCGCGGCAACCCCTACTGTGCTCCGCTGAGCGGG-----GC 2412
QY 2087 gProLysGlnProGly-----ProValLysLeuGl 2097
Db 2411 GCCCGCTCGCGCGCGCTGGACTCTGACTCTGACTCTCGGCGCCCTCCACCGCT 2352
QY 2097 yGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSe 2117
Db 2351 CCTCGAGCGCTTCCTCTCGCTCTTCTCTCCCTCGCTCTCGCGCGGCTCTCGTCTCG 2292
QY 2117 rProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg-ValValThrLeuAlaG 2137
Db 2291 TCCTGTCGCCGCTCTCTCTCTGCGTTCGCGCGCGCGCGCTCGGTCCTCGC- 2233
QY 2137 lnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerA 2157
Db 2232 ----- 2217
QY 2157 laProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuA 2177
Db 2216 CCGCGCGCGCTGCGCTCCCGCGCGCGGTGGCGCTCTCTCTCTCTCTCT- 2170
QY 2177 rArgProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaArg-Gly 2196
Db 2169 -----CCTTCTCTCGCGCGGATATCCCGCGCGGAGGT 2136

QY 2197 SerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216
Db 2135 GCGTGGCGCGCGCGGAGGTGCGCGCGTGGTGAAGCGGC-----GGC 2091
QY 2217 GlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis 2236
Db 2090 GCGCGCGCGCGCGGA-----GGGCTCGCGCGCGGAGATCG 2055
QY 2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256
Db 2054 TCCCGGTCCCTCTCTCTCCGCGGTCCCGCG-----TCCCGTCTCTCTCTCTCC 2001
QY 2257 MetGly-----SerLysSerProGlyAsnThrSer 2266
Db 2000 ATCGGCTGAGAAAAGAGTTATTTTTCAGAGTCAGAAAATAAAGTTTGTGCTGATTTTCT 1941
QY 2267 GlnPro-Pro-----AlaPheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
Db 1940 GAACCAAGCCGAGTCTCTGAGATTTTTTGGGAGATGGAGCGGCCATCTTGGCGGTGT 1881
QY 2283 lLysSerLysGlnGlnIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGl 2303
Db 1880 CTCTGGGTGGAGGTGCTCTTGTGGATGGGGTC-----CCTGGTGGGAGGAAGAAGAG 1824
QY 2303 uTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly----- 2321
Db 1823 GTGAGGCTTGTGGTGGGTGACGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1764
QY 2322 -----ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSe 2337
Db 1763 TGGGAAGAAGTGGATGGGGTCT 1704
QY 2337 rThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLys----- 2353
Db 1703 TCAGCAGATGGGGTC-----CTCCTCTCTGGAGGTCTTGTGGGAGAGTAGAGGT 1650
QY 2353 ----- 2353
Db 1649 CTTGGGATGTTGGGGGTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1590
QY 2353 ----- 2353
Db 1589 GGTCCCGCGGGTCTCTGTGGGATGTAGATGTCCGAGGGTCCCGGTGTCTCCGGGTGG 1530
QY 2354 -----TyrAspGlnTyrGluGluSerProPr 2362
Db 1529 GAGTTAGACGATGGTGGTCCGTATGCGACTGCCCATCT- GCCTGGACGTCCGCGCAC 1471
QY 2362 o-----LeuSerAlaAsnAlaPheAsnPr 2370
Db 1470 GAGCGCAGACGCTGCGGTGCATGCACAAAGTTCTGTCTGAGACTGCATCCAGCGCTGAC 1411
QY 2370 oLeuAsn-----AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
Db 1410 CTGACAGACACGCTGCGCGCTGTGCAATGCCCGGTCCACCTCA----- 1365
QY 2387 pGlyArgSerAspHisThrLeuThr----- 2395
Db 1364 -----TCTCCACCACGTGGACGACGCTTCTGTTGTGGAGCGCCCGTGGAGGG 1312
QY 2396 -----SerProGlyGly-----GlyGlyLysAlaLysValSerGl 2407
Db 1311 GCGACGATGTCACGCGGAAGAGGATGAGCCGTAGGGGGGGTTCGCGGTCTCATCTGG 1252
QY 2407 y-----ArgProSerSerArgLysAlaLysSerProAlaProGlyLe 2421
Db 1251 GGTGAAGACTATACCGAGGAGGTGCCACGAGGAGCGGAGGGGAGGGGCTCCGGGTCT 1192
QY 2421 uAlaSerGlyAspArgProProSerValSerSerValHisSerGluGly-----As 2438
Db 1191 GGTCTCGCGCGCGCGCGAGAGTCCCGCTCTTCACTGGCTTTATGGGCGAGGTCTCGACG 1132

1292 luAspGlyArgSerSerSerGlyPro-----ProHisGluThrAlaAlaProLysArgT 1310
: : : : :
5052 AACCAAGTTCCAGCGCGGTCACAGCGCCACAGCGCCCTCTTCATCACCGGC 4993
1310 hrTyRspMet-----MetGluGlyArgValGlyArgAlaIleS 1323
4992 AGCGTGACCCCGCGCTGCGGCACATCGGGAGCCATGCGCCCGCAGGACCGCTCTGG 4933
1323 erSerAlaSerIleGluGlyLeuMetGlyAlaIleProGluArgHisSerProH 1343
4932 GCCCTGCGCACGCGGTGAGCGCGTGGCCATGAG--CGCGCTACGATCGACCCAG 4876
1343 iHisLeuLysGluGlnHisIleArgGlySerIleThrGlnGlyIlePro----- 1360
4875 AAGACCTTCACTCTGCAGACCTCCCGCGGCGCTACGCGACATGCGCTACCCGGGCGC 4816
1361 -----ArgSerTyRValGluAlaGlnGluAspT 1370
4815 GCGCGGACCCCGCGCGGGAGGCGACGTCGAGGCCCTCGCGCCGCTCCGCGCC 4756
1370 yrLeuArgArgGluAlaLysLeuLys-----ArgGluG 1382
4755 GCCTTCGCGCGCGCAGCGCGCGTGCCTCGCGGAGTGCGGACGCGCTCGCTCG 4696
1382 lyThrPro-----ProProProSerArgAspLeuThrGluAla 1396
4695 GCCTGCGCGCGGCTCTGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4636
1397 TyRlysThrGlnAlaLeuGlyPro-----LeuLys 1406
4635 CGCGCCCGCGCGCTCTGCGGCGCGCTGCTGCTGCGAGGAGTGACCGCGCTGCTGCG 4576
1407 LeuLysProAlaHisGluGly----- 1413
4575 CTGCGCAGCGATCCCGCGGCGCGCGCGCGCGAGCGAGCGCGCGGACTGTGTC 4516
1413 ----- 1413
4515 GCCCTGTCGCGCGACCGTGGCGCGGTGTGCGCTACAGCTGAGCGGGCGCGCGCC 4456
1414 -----LeuValAlaThrValLysGluAlaGlyArg 1423
4455 CGCAGGCGCGCTGACCTACGCGCGCGCTCTTCGCGCGCGCAACGTCGCGCGCGCC 4396
1424 SerIleHisGluIle-----ProArgGluGluLeuArgHisThrProGlu 1438
4395 CGGCTCGCGAGCG 4345
1439 LeuPro-----LeuAlaProArgProLeuLys 1447
4344 CTGCCCCCGCTCTGCGCGCAGCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 4285
1448 GluGlySerIleThr-----GlnGlyThrProLeuLysTyRAspThr 1461
4284 GCGGCGCGCTCTGCGCGCTCTGCGCGCTCTGCGCGCTCTGCGCGCTCTGCGCGCT 4225
1462 GlyAlaSerThrThrGlySerLysHisAspValArgSerLeuIleGlySerProGly 1481
4224 GGCTCCAGCACAAGTCCAGCTC-CGGCACAAGTCCGCGCTCTGCG-GCTCTCTCGGCT 4167
1482 ArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu-Ar 1501
4166 AGCCAGCTCCCGCG-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4128
1501 gAlaCysTyRThrGluGluSerLeuLysSerArg---ProGlyThrAlaSerSerGlyG 1520
4127 AGCGCGCAAGAAAGCG 4068
1520 ySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPr 1540
4067 A-----CGAGGGCTCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCT 4020
1540 oLeuThrTyRArgHisGlyAlaProPheAlaGlyHisLeuProArgGly-SerProV 1560

4019 GCGACGACGAAGAGGACCGGGCGCGCGCGCAAGCGCGCTCTCTCGACTCGGGCGCG 3960
1560 alThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAla 1580
3959 CTCCG-----GACCCCGCTCCCGCGCTGCTCTCTCTCTCTCTCTCTCTCTCT 3906
1580 erGluAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrV 1600
3905 AGGACGACCG-----CTCGGCGCGCGCTCGGGCCCA 3873
1600 alProGluHisHisProHisProIleSerProTyRHisLeuLeuArgGlyValSerG 1620
3872 TGCCCGAG--CACCCCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3823
1620 lyValAspLeuTyRArgSerHisIleProLeuAlaPheAspProThrSerIleProArg 1640
3822 -----GAGACCCACACCGCGCG- 3805
1640 lyIleProLeuAspAlaAlaAlaTyRLeuPro-----ArgHisLeuA 1656
3804 -----CCCTCCCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3750
1656 laProAsnProThrTyRHisLeuTyRProProTyRLeuIleArgGlyTyRProAspT 1676
3749 TGGACCGAGGCTCTCCCGAACTTGGCGCGCGCGCTCTCTCTCTCTCTCTCTCT 3699
1676 hrAlaAlaLeuGluAsnArgGlnThrIleAsnAspTyRLeuThrSerGlnGlnMetH 1696
3698 CGCGCGCGCT 3667
1696 iHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProA 1716
3666 --GGGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3609
1716 rgGluSerSerLeuAlaLeuAsnTyRAlaAlaGlyProArgGlyIleIleAspLeuSerG 1736
3608 CCAGAGAC-----GTGCGCGTGGTGTCTCTACGAC 3576
1736 lnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetA 1756
3575 CGTGTCCCGCAGGAGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3516
1756 sp-ArgLeuAlaTyRLeu-----ProThrAlaProGlnPro 1767
3515 ACCCGCGCGCGCGCT 3465
1768 PheSerSerArgHisSerSerPro-----LeuSerProGlyGlyProThrHis 1784
3464 ---TCTGACGCGGACTCGACGCTTGGCGCGGAACTGGACCGGCGCGCGCGCATCG 3408
1785 Leu-----ThrLysProThr 1789
3407 GCGCGCTCAACGCCAGGGGTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3348
1790 ThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgGlu 1809
3347 CCGTCGAGTACCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3309
1810 ArgGluLysSerIleLeuThr-----Ser 1817
3308 GCGCGCTCATCTGCTGTGGACCATCGAGACTGGCGCGGACGCGCGCGCGCGCG 3249
1818 ThrThrThrValGluHisAlaProIleTyr---ArgProGlyThrGluGlnSerSerGly 1836
3248 ACTACCAAGTACGTCGCGCGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCG 3189
1837 SerSerGlySerSerGlyGly-----GlyGlyGlySerSerSerArgPro 1851
3188 CCGAGTCCCGGAGCTCCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3129
1852 AlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeu 1871

Db 7054 CAGGTACGACACATCGTTCTCTGCCAACCCGAGGGGATCCGATCGTCTCCGCTCCGGC 6995
QY 735 -----GluCysSerGly-----ProAlaThrValAsnAsnSerSerAsp----- 747
Db 6994 GCGGACTCTGAGACTCCGGCTCTCCGGCGGCTATCAGCCC-TCGAGGAGGCCGACCC 6936
QY 748 ThrGluSerIle-----ProSerPro 754
Db 6935 ACAGAGCTCTCGGCCCCACCCAGAGAAGAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 6876
QY 755 HisThrGluAlaAlaLysAspThrGlyGlnAsnGlyPro----- 767
Db 6875 TTCTTCTCCGCGCCCGC-----TCTCCGCGCTCGGCGCCCGCTCGCTCAGGCAGAAAGA 6819
QY 768 -----LysProAlaThrLeu 773
Db 6818 CCCCAGATCGAGACCATGGCCGACGATCTCTTTGACTTTCAGAGACGAGGGCACTTCA 6759
QY 774 GlyAlaAspGlyProProGlyProThrProArgArgThrSerArgAlaPro 793
Db 6758 GCCAGCTCTCGCGCGCCCGC-----CGCCGCGCCGAGGAAGAGGATCGCTCGCGAGATCTCC 6705
QY 794 IleGluProThrProAlaSerGluAlaThrGlyAlaProThrProPro----- 810
Db 6704 ---GCCCCGACGCGCGCAGCAGGCTCCCGCGCCGCGCTCTCTCGCGAGATCTCC 6648
QY 811 -----AlaProProSerProSerAlaProProValProLysGluGlu 826
Db 6647 TCTTCGCGCGCGCGCTCTTCTCGACACGCGCGCGAGGCGCGCTCTCTCG 6588
QY 827 LysGluGluGluThrAlaAlaAlaProProValGluGluGluGluGluGluGlu 846
Db 6587 CGCGCGCGCGCGA-GCCACGAGCGCGCC-----CGTCTCTCT 6550
QY 847 AlaAlaGluLeuAla-----ValAspThrGlyLysAlaGluGluProValLysSerGlu 885
Db 6549 TCGGCCACGACGAGCGCGCACGCCCGCGGCTCCGCGCGCGAGATCGTGTGCTAGACGAC 6490
QY 866 CysThrGluGluAlaGluGlu-----GlyProAlaLysGlyLysAspAlaGluAlaAla 883
Db 6489 GAAGACGAGGAGGAGGAGCGCGGTCCCGCGCGCGCGGTCCCGCGTCCGGGTCTC 6430
QY 884 GluAlaThrAlaGluGlyAla-----LeuLysAlaGluLysLysGluGlyGlySerGly 901
Db 6429 CATCAGGGCTCCGAGCAGCTACCTCGTCTCGGCGCGCGCTCCCGCGCGCGCTCCGG 6370
QY 902 -----ArgAlaThrAlaLysSerSerGlyAlaProGlnAspSer 915
Db 6369 CCCCAGCGCCGACCCCGGCGAGCTCGACGCGGACAGCGCGGCGCCCGCGCGCC 6310
QY 916 AspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly----- 931
Db 6309 GGGGCTCTGCTCCCTCCGAGCTCCCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6250
QY 932 ---GlyAspLysAsnArgLeuLeuSerProArg-----ProSerLeuLeuThrPro 947
Db 6249 TCGGCTCCCGCGCGCTCCCGCGCGCGCGGAGATGGTCCCGCGAGAGGCGCGGACCCG 6190
QY 948 ThrGlyAspPro----- 951
Db 6189 GTCGGTGAGCCCGGCGCGCGCGCTCCGACCCCGCGCGCGCGCGCGCGCGCGCGCG 6130
QY 952 -----ArgAlaAsnAlaSerProGlnLysProLeu-AspLe 963
Db 6129 GCGGTGGCGCGCGCGCGCGCGCGCGCTTCCCGAGCTTCCCGCGCGCGCGCGCG 6070
QY 963 uLysGlnLeuLysGlnArgAlaAlaAlaProProIleGlnValThrLysVal----- 981
Db 6069 CCGTCTCGCGCGCGCGCGCGCGCGCTCCCGCGCGAGGAGGACCGCGTCTCGTAC 6010
QY 982 -----HisGluProProArg----- 986
Db 6009 CATCACGACGCGCGAGCCCTCTCTCGACGAGCCCGCGCGCGCGCGCGCGCTCTCT 5950

QY 987 -----GluAspAlaAlaProThrLysProAlaPro-----ProAlaProPr 1000
Db 5949 CGGCCGCTGGCGCGCGCGAGCCCGCTCTCTCCAAACCCCACTCCAGCTCCAGCTCCACC 5890
QY 1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGln----- 1014
Db 5889 ACCACGCTCGCGCTCGAGCCCGTCCGCGCGCGCGAGAAAGGAGGAGGAGCGACTCGGC 5830
QY 1015 ---ProGlySer-----SerProArgGlyLysArgSerPr 1026
Db 5829 CTCGCGCGGAGCGCGCGCGCGCGCTGACAGACAGCCCGCGCGCGCGCGCGAGAG 5770
QY 1026 oAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeu----- 1042
Db 5769 GCGCCCTCCGCGCGCGCGCTTCAGCAGCTCCAGCAGCGCGCGCGCGACTCGGAC 5710
QY 1043 -----ProGlyAspProProCysTrpThrSerGlyLe 1053
Db 5709 CTCTCCCGCGCGCTCGCCCTCCGCGCGCGCGCGCGCGCG----- 5667
QY 1053 uProPheProValPro-----ProArgGluValIleLysAl 1065
Db 5666 -CGCGCGCGCGCGCGCGCGCGCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5608
QY 1065 aSerProHisAlaPro-----AspProSerAlaPheSe 1076
Db 5607 TCGTCTCTCTCTCTCTCCGAGGAGAGAGAGAGAGTCCGCGCGCGCGCGCGCTC 5548
QY 1076 rTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLe 1096
Db 5547 GCGCGCGCGCGCGCGCGCTCTC-----CGCGCGCGCGCGCGCG 5506
QY 1096 uProArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSe 1116
Db 5505 GCGCGCGCGCGCTCGCGCTCTCCGCTCCGCGCTCCGCGCGCGCGCTCCCG 5446
QY 1116 rValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValPr 1136
Db 5445 GCGCGCGCGCGCGAGC-----CC 5428
QY 1136 oTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMe 1156
Db 5427 GCGCGCGCGCGCGAGAGCGCGCTCCACCAACACACCTCTCTCTCTCTCTCTCTCT 5368
QY 1156 tAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArg-- 1175
Db 5367 GGGCCCCCGCGA-----CGACGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5329
QY 1176 -GlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlnAlaSerValLe 1195
Db 5328 GGGAGCCCTGGCGCGCTCCGACCTCCG--CGACGCGCGCTCTCTCTCTCTCTCTCTCT 5272
QY 1195 u-ArgGlyThrAlaLeuGlySerValProGlySerIleThrLysGlyIleProSerT 1215
Db 5271 GCGGGGA-----CTCCGCGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5236
QY 1215 hrArgValProSerAspSerAlaIleThrTyArgGly-----SerIleThrHisG 1232
Db 5235 GAGTCCGCGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTCTCTCTCT 5177
QY 1232 lyThrProAlaAspValLeuTyLysGlyThrIleThrArgIleIleGlyGluAspSerP 1252
Db 5176 ----CGCGAGAT-----GGGGACTCGAGGAAGCAGCAGCGCGCTCTCTCTCTCT 5134
QY 1252 roSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyGluG 1272
Db 5133 CGCTCATCTACAGCGCGCGC----- 5111
QY 1272 lyLysLysGlyHisValLeuSerTyGluGlyGlyMetSerValThrGlnCysSerLysG 1292
Db 5110 --CGCGAGGCCATGCTCTGGTCTGAGAACCCCGCGATGCGGCCCGCGCGCGCTCT 5053

[illegible]

1271 luGlyLysLysGlyHisVal-LeuSerTyrGluGly-----GlyMetSerVal 1286
11418 CAGCGAACAACGCGCAACATAGCGTCTCGCGCGGATTCACCTGCGTAGCTCGCATGCC 11359
1287 ThrGlnCysSer-----LysGluAspGlyArgSerSerSerGlyPro 1300
11358 TCGCACTGTAGTCACGACGCGCGGTAGAAAATCGAGTGGCGAAGCCGCGCAGTGGC 11299
1301 ProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGly--- 1319
11298 CCTCGGTGAACCATCTCCGAGTAGCGCTTGTCCACAGCAAGAACCCGACAAAGCGTT 11239
1320 -----ArgAlaIleSer---SerAlaSerIleGluGlyLeuMetGlyArgAlaIle 1335
11238 GCTCGCGGAGGTGCTATCACCAAAATCGGGTGGGTGGCTTGGCTGAGGTGTACAGTTT 11179
1336 Pro-----ProGluArgHisSerProHisHisLeuLysGluGlnHis 1349
11178 CCGAGATACCTCGACGCTACCGGTCGACGAGTTCTCCGCGGTGCGCCCGTTGGCGA 11119
1350 HisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAsp 1369
11118 -----GTTCTTGTCTCAACACACGCGCGCACAG----- 11092
1370 TyrLeuArgArgGluAlaLysLeuLeuLysArg-----GluGlyThr 1383
11091 -----CCTCACGATCTCGCGGCGCGCGCGTACGACAGCAACGAGTGTACATGAACG 11038
1384 ProProPro-----ProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
11037 AGGCACCTCCGGGTGATTCCACCGCACCGC-----GCA 11002
1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGlu 1420
11001 GCCCGCGGCGG----- 10990
1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
10989 -----GTTCTCGC-----CGATCAACCCACAGAT----- 10966
1441 LeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAsp 1460
10965 -----CCCCACCGTCCGCACATCCAGTGG-ITGGCGCGTGGCGAGATCTCTCCACG 10913
1461 Thr-----GlyAlaSerThrThrGlySerLys----- 1469
10912 ACATCGGTGATGATCGATGAGTCTGCAAGCTGCTCGGATCGCTGCGAGGTTCG 10853
1470 -----LysHisAspValArg 1474
10852 GTGGACAGCAGATAGACGTGGCCAGTTCGATCGCGGCTTCGTGGCACAC-----CGC 10799
1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
10798 AGCAATTC-GGGGATCTTGGCGCACCCAT---CCGTTAGCGGTAGCTGACGTGCTCGTA 10743
1495 AspAlaArgAlaLeuGluAlaCysTyrGluGluSerLeuLysSerArgProGlyThr 1514
10742 GCCCGCGTGGCGGCCATCGCGGTGCGGTGCGCAGCAGCGGTATGTGCGCGGGCAG 10683
1515 AlaSerSerSerGlySerIleAlaArgGlyAlaProValIleValPro-----Glu 1532
10682 GTCCGATTCGAGCGGCGCAAGCCCTCGCGCAGCGCAGCTCGTAGCGCGGTACACGG 10623
1533 LeuGlyLysProArgGlnSerProLeu---ThrTyrGluAspHisGlyAlaPro-PheAl 1551
10622 CTCCTTGGCGCGGGGATGATCTCCACAGATACCCAGACTACTGTGACCCAGACGTC 10563
1551 aGlyHisLeu-ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnG 1571
10562 GAGTCAGATGTCGCTACCTGGCCCGGACTACTGTGTTCCAGCACACCCACCGCGCAGCAC 10503

1571 luGlySerLeuSerSerSerTyrAlaSerGlnAspArgLysLeuThrSerThrProArgG 1591
10502 AAGACGAGCAACAGCGGAATGAGCGGCAG-----GCCG 10467
1591 luIleAlaLysSerProHisSerThrValProGluHisProHisProIleSerProT 1611
10466 ACACCGCCACACCGCGGAAGCAAGACCGCGCCACGCTGCCAC----- 10420
1611 yrGluHisLeuLeuArgGlyValSerGlyVal-----AspLeuTyrArgSerHisI 1628
10419 -----CACCTGTCGAGGTGTGCGCGGTCTCTGACCAACCGCGCTTCGCGGTGGA 10365
1628 leProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAla- 1647
10364 TCCACGTGTATTCCGCGCGCACCGCGGTGTGCGCGTGTGCGGTGTGCGTGTGCGT 10305
1648 AlaTyrTyrLeuPro-----ArgHisLeuAlaProAsnProThrTyr 1661
10304 GCGCGGTGGTTCGCGCAAGCGGCGCTGACGAGCTGGCCTACACCGCGGCACGATCA 10245
1662 ProHisLeuTyrProProTyrLeuIleArg-----GlyTyrPro-AspThrAlaAl 1678
10244 CCATGTTACCGTCAGCGCCACTTATCACCGGCTCAATTGGAATCCGCGCAGCGCCGCA 10185
1678 aLeuGluAsnArgGlnThrIleIleAsnAspTyr---IleThrSerGlnGlnMetHisH 1697
10184 ACTGGATGAAGCGCGCGCACCACTCGATGATCTTCTGTTTCATCGCGCAGCTACAC-- 10127
1697 sAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgG 1717
10126 -----GCCGTCGCGTGTGCTGCTTGCCTTGCCTTGCCTGCGC-----CCACGACG 10092
1717 uSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp----- 1733
10091 GCGC-----CGTGTGTGTGCTGATTTGTTGGGCG 10062
1734 -----LeuSerGlnVa 1737
10061 GTCGCATCGCGGAATCTTCTGAAGATGTGTTGGCGCGCGCGCGCTCGGTGGGG 10002
1737 lProHisLeuProVal-----LeuValProPro----- 1746
10001 TTCGCTGTACTTGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 9943
1747 -ThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProG 1766
9942 AACGCGCGGTGACCGCGTGTGCTGCTGTTGTTGGCGCGCTGTGTACAGCATCGGC 9883
1766 nProPheSerSerArgHisSerSerProLeuSerProGlyGlyProThr-----Hi 1784
9882 GGCATCTCTACGCGGTACGCT-----GGCCCGACCGGTGGCCG 9844
1784 sLeuThrLysProThrThrSerSerSerGluArgGluArgAspArg-AspArgGluA 1804
9843 ACNAGCTTCGGTACACAGAGTCTTCCACGCTGCACCGCGGTGGCGGTGGTGGTGGTGG 9784
1804 rgAspArgAsp-----ArgGluArgG 1811
9783 TACATCGCATGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 9724
1811 lLysSerIleLeuThr-----SerThrThrValGluHisAla---P 1825
9723 AAAATCGCAGCGCAGCAGCGCAGCAGCGCGGTGGGATTTACGCTGGAGCGGTGCCACC 9664
1825 rolleTtpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGly----- 1842
9663 CGAGTCCGTAGCGCGGTACGCCCAAAGCGGTGGCGAGTTCGCGCGGTGTGTACCGGC 9604
1843 -----GlyGlyGlyGlySerSer-----S 1849
9603 AGATCGCAGACCCCGCGCGCGCATACGTAGCGCGGTGGCGCGCGG-CCACCCGATCCCG 9545
1849 erArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnA 1869

WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Alignment Scores:
Pred. No.: 1.09e-09 Length: 110000
Score: 618.00 Matches: 646
Percent Similarity: 30.90% Conservative: 267
Best Local Similarity: 21.86% Mismatches: 1041
Query Match: 4.68% Indels: 1009
DB: 4 Gaps: 132

US-09-522-753-5 (1-2517) x AAI99682_12 (1-110000)

Qy 150 ProProSerProProHisThrAspProGluLeuGluLeuValProProArgLeuSerLys 169
Db 15050 CCACCATCCCATCGGAGCGACCGCATACGCTCGCGCTATTCGCCCATCGACCCCGGTG 14991

Qy 170 GluGluLeuLeuGlnAsnMetAspArgValAspArgGluLeuThrMetValGluGln 189
Db 14990 GAACACGGCTC-----GGCTTGGCGTGGTCATGATGGCG 14955

Qy 190 IleSerLysLeuLysLysLysGlnGlnGlnGlu 202
Db 14954 CTGTGCGCCCTATCGCGCATGCTGCTGTGTGCGAAATTCGGGTTGGCGCGCGACCC 14895

Qy 203 GluAlaAlaLysProProGluProGluLysProVal 214
Db 14894 GAAGTGGCGCTGTGCGGAGCGCGCATTCGCTCCCGTCACCTGACCGGCGACGGCGTTGACG 14835

Qy 214 ----- 214
Db 14834 GCGCCCCGCACTGCGCGCGCATGCTCTGAGCTGTGACGATTGCGGCGCTCGCGCGCG 14775

Qy 215 -----SerProProLleGluSerLysHisArgSerLeuValGluIleTyr 231
Db 14774 GCGAAAGACTCGCACCCCGCCAGTCAAGTGGCGAC----- 14739

Qy 232 AspGluAsnArgLysLysAlaGluAlaHisArgIleLeuGluGly-----Leu 248
Db 14738 -----AAACAGGGTTGAAACCGCGCCGACCTGCGCGTGGAGGGGTGATATTCCTC 14688

Qy 249 GlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyr----- 266
Db 14687 GCGGTTTCGTTGAA-----CAGCTAGCGATGGCGCGCAACACTTCATC 14643

Qy 267 -----HisGluAsnIleLys-----IleAsnGlnAlaMet 276
Db 14642 GCCACCGCGCGCACACTCGCGTTGTGCGCGCGACCGCGCGGCTTCGCAACGCTAAC 14583

Qy 277 ArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyr----- 294
Db 14582 GCGCGAGCGCAATCCGCTGATATCCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14526

Qy 295 -----LysGlnLysPheCysGln-----ArgTyr----- 302
Db 14525 ATCATGTATGATATCCGCAATCTTTACCCACATATAACGGTTTATATATAAATCATC 14466

Qy 303 -----AspGlnLeuMetGluAlaLeuLysLysValGluArgIleGluAsn 318
Db 14465 ATAAGCACTTACAGCCGATGATACGTTCTTATCATCGATGATATCATTTGTTGACAACG 14406

Qy 319 AsnProArgArgArgAlaLysGluSerLysVal-ArgGluTyr-----Ty 333
Db 14405 GACCCCT-----ATAACGGTTTATCTACTGTACACTAATTT 14370

Qy 333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerAr 353
Db 14369 CCATAAGCTTTTAGATTTCGACGCGCCGAGCAATGATCGATCGAAACCGAATCAACCGCC 14310

Qy 353 gValGlyGlnArgGlySer----- 359
Db 14309 AGCCGGGAGCGGTGATCCCGTTCGATAGGCGCGGGAACGTAGGGGATCGACGCG 14250

Qy 360 ----GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleLeuSerG 378
Db 14249 TGTGGCGTACTGTAGTCACGCGCGCGCAGAGTCCAGCTCGCG----- 14204

Qy 378 yLeuSerGluGlnGluAsnLeuLysGlnMetArgGlnLeuAla----- 393
Db 14203 -----AAACTCCCGCCAGTGGGCGCTCGGTGGCGCGGAC 14172

Qy 394 ----ValIleProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMe 412
Db 14171 AAGCGTTGTTCCCGAGGTGCGGATCACAGGTGCGGATCGGGT----- 14126

Qy 412 tAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTr 432
Db 14126 ----- 14126

Qy 432 pSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheG 452
Db 14125 ----GCCCGCAGGTATTTCAGGTTTTCGAGATGCTTCTACGCCACCG----- 14081

Qy 452 yLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLe 472
Db 14081 ----- 14081

Qy 472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgSerTyrArg-----ArgAr 490
Db 14080 -----CGTCGACGAATTCGACGCGTCCCG 14055

Qy 490 g-GlyLysSerGlnGlnGlnGlnGlnGlnGln-----GlnGlnGlnG 507
Db 14054 TAGCCCAACGCGCAAGCTCGTGGGATCGCACAGGTTTCTGTGGACAGCGGCGAG 13995

Qy 507 lGlnGlnGln-ProMetProArgSer-----SerGlnGluGluLysAsp 521
Db 13994 ACGTAAACAGTCCGATGCTCATCAGCGGTACCGCTCGCGCAGCTGTAGTACGCGACT 13935

Qy 522 GluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAsp 541
Db 13934 CAGACCGGACTGGATTTCGCGCGCACGCGGACGCGGACCGCGCT----- 13890

Qy 542 LysGluAspLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLys 561
Db 13889 -----CGCCGCGATCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13854

Qy 562 GluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArg 581
Db 13853 -----TCGCCCCCGGCTCGCGACAGCAGCCGCTGAGCGCGCGCGCGCGCGCGCG 13806

Qy 582 IleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSer 601
Db 13805 CTGACCGGTTG-----GCACCGCGGTGCGCG 13779

Qy 602 AlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMetGlu 621
Db 13778 CGGCCCCCGGCGCGCGCTGCGGATCGATCGGCGTTCGCGCGTGCACCGCGCGCGCGCG 13719

Qy 622 ThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetVal 641
Db 13718 CACCGCACCGGCTGCTGCTGGGTGGGCGCGCGT-----CGCGGAACA 13674

Qy 642 GlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsn 661
Db 13673 GCCAGGTTCGCGCGCGTTCG 13614

Qy 662 LeuAspGluIleLeuGln-----GlnHisLysLeuLysMetGluLys 675
Db 13613 CGAAGTTCG 13554

Qy 676 GluArgAsnAlaArgArgLysLysLysAlaProAlaAlaAlaSerGluGluAlaAla 695
Db 13553 CACCACCGCGCACCGCGTCCC--CAACAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13496

Qy 696 PheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGlu 715

```
Db 15279 GTAGGCGACGTCTCCAGGTCCTGGTGGGGTGCTCGGTAGGTGCGCGCGCGACGCCGCGC 15220
Qy 2145 -----GlnAspTyrThrArgHisHisProGln 2153
Db 15219 CGCCCGGGCGCGCAACCGCTCGCGGTGCTCGCGGAGACGACGACGACCGCGGCC 15160
Qy 2154 Gln-LeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer----- 2170
Db 15159 CATCCGGTCCGGTCCGGTGTGTCACCGGTC-----GGCGGGCGCTCTCTCGATGAT 15109
Qy 2171 ---CysProValLeuAspLeuArgProProSerAspLeuTyrLeuProProAs 2189
Db 15108 CGCGTGTGCTGCTGCTGCGCGTGATG-----CCGAA 15079
Qy 2189 pHisGlyAlaProAlaArg-----GlySerProHisSerGluGlyGly-----Ly 2204
Db 15078 CGAGGACACCCCGCGCGCGCGCGCTCGCGCGGGCGCGGCGCGGCTCTCTGGAG 15019
Qy 2204 sArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyAspGly----- 2221
Db 15018 CAGGCTGATCGCGCGCGGACGACGATCGATCTCCGGTACGCGCTCGTGGCGGTG 14959
Qy 2222 -----IleGluProValSerProPr 2228
Db 14958 GCGGGCGAGGACACCGCGCTGCATCGCCACACACAGTTTCATCAGACCGCGGACCG-- 14901
Qy 2228 oGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAs 2248
Db 14900 -----GCGGGCGCTGGGTGTGCTGCCGATGTTG----- 14874
Qy 2248 pGlyGluGlnThrGluProSerArgMetGly-----SerLysSerProGlyAsnThrSerGl 2267
Db 14873 ----GACTTCACCGAGCGCATCGCGAGCGGTCTCACGGTCACCGCGTACGCGTCGGA 14818
Qy 2267 nProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVal----- 2283
Db 14817 GAGCGCCCGTGC-CTCGATCGGATCACCGAGCGGGTGGCGGTGCGGTGCGGTCTCCACGA 14759
Qy 2284 -----LysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAs 2300
Db 14758 CGTCAGCGGTGGGGCGTCAGCCCGGACTCGCGGAGGGCGCTCTCGGATCAC----- 14707
Qy 2300 nGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleTh 2320
Db 14706 ----CCGACCTGGGC-----CCGCGCGCTTCG 14684
Qy 2320 xGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMe 2340
Db 14683 GGGC-----GGAGATCGCTGCTCGCGCGGT 14657
Qy 2340 tGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnTyrPgluGluse 2360
Db 14656 CGGAGTTGACAGC-----GGTGGCCCTGATCACGGCGAGCACCG 14618
Qy 2360 rProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerLeu----- 2377
Db 14617 GGTGGCGGTGGCGGGCGTCCGAGCGGTTCGA-GCAGGAGCATCCCTCGCGCTTCG 14559
Qy 2377 ----- 2377
Db 14558 GCATGCCGAACCGTTCGGCGGGCGCGAGAACCCCTTGACCGTTCGTCGGGGCGAGG 14499
Qy 2378 -----ProAlaMetProIleThrAlaAlaAspGlyArgSerAs 2391
Db 14498 GAGTTTCATCCGACTGAAGTCCACGAGCATCCCGGGTGTCCGATCACCGTCAGCCACCG 14439
Qy 2391 pHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSe 2411
Db 14438 GCGA-----GCGCCATCGTCGACTCCGCGCGCGGACGACGTGGCACCGCATGTC 14388
Qy 2411 xArgLysAlaLys-----SerProAlaPro----- 2419
Db 14387 ACGCCGACGCGGACGACGACGAGCGGGGTGTCAGCTGATCGCGCGCGCTCCAGCGCG 14328
```

```
Qy 2420 -GlyLeuAlaSerGlyAspArg----- 2426
Db 14327 AGGGTGTAGGCGACCGACCGGAGCGCGCTGGTGGTGGTCCCGGTCTCATCAGGTAGCCC 14268
Qy 2427 -----ProProSerValSerSerValHisSerGluGlyAspCysAsnArgAr 2442
Db 14267 TCGACGCCCTCACCCCTCGGCCGACCGCGGGGCGGTACTCTCTGGGGGATCAGACCGACG 14208
Qy 2442 gThrProLeuThrAsnArgValTyrPgluAspArgProSerSerAlaGlySerThrProPh 2462
Db 14207 AACACCCCGTTCGGGAGG-----TCCGCAACGACGTCCGGGGGATCCCGGCGCGT 14157
Qy 2462 eProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProPr 2482
Db 14156 TCCAGCACCTCCACGACGACGCTCCACGTCGATCCGCTGCTGCGGCTCGACGGCGAGTGC 14097
Qy 2482 oGlyLeu----- 2484
Db 14096 TCCCGTGGCGACAGCCCGAAGAGGAGCGGTGCAAGAGGTGGCGCGGTGAGGAAGCCA 14037
Qy 2485 -----ProAlaGlySerGlyPro 2490
Db 14036 CCAGCGCGTGGTGGCGCGTGGCCGACCGGTCGGGTCCG 13997
RESULT 70
AAI99682_12/c
Continuation (13 of 45) of AAI99682 from base 1200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
```


Db 19264 -----GACCCA----- 19259
Qy ProPheProValProArgGluValIleLeuAlaSerProHisAlaProAspPro--- 1072
Db CCCACCCGACCCACCCATACGTCCCAACACGCCCCCACTCCACCGGATCCCCA 19199
Qy SerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHis 1089
Db ACAGCGTCCCGTCCCATGGCGCTCCACACACCCACATCCACCCGACACCCGAC 19139
Qy 1090 -----AspThrAlaArgProVal--- 1095
Db GACCCACGCCCGCGGATACACCGGTGTCGGCCACCCCGGCGCGCCGACCCAT 19079
Qy 1096 ---LeuProArgProProThrIleSerAsnProProLeuIleSerSerAlaHis 1114
Db TACTCGCCCATCTGATTCCACCGCGAACCACCCACACACC-----CAACA 19031
Qy 1114 sProSerValLeu-----GluArgGln----- 1121
Db CGCGACGCCCTCCCGCACCGCCACACGACCAACCGCTGCAACACGACGACCGACCCCT 18971
Qy 1122 -----IleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValPr 1136
Db CCCCCAGACGACCGCTCGGCTGTCGAGAGGGCTTGCACCTCGCTCCGGAG-CC 18912
Qy 1136 oTyrSerGluHisAlaAlaAlaProValGlyPro----- 1147
Db AACCGGCCCTCGCGGAGAACTCCCTGAACACCTCCGGACCGGCCATCACCGACACCCA 18852
Qy 1147 ----- 1147
Db 18851 CCGCCACCGCCAGACACAGTCCCCAGAACGAAACCCCGCGCACGTGCAACGCC 18792
Qy 1148 ---ValThrMetGlyLeuProLeuProMet-AspProLysLysLeuAlaPro----- 1163
Db ACAAGCACGACGAAACAGCGGTGTCCAGTATCCCGGCCCTCCAAACCCCAACAG 18732
Qy 1164 -----PheSerGlyValLysGlnGluGln----- 1171
Db TACCGATCCGACCGGAGCGAGCTCGCGGTGTTCCTGTCAACAGGTAGCGCTCGACC 18672
Qy 1172 ---LeuSerProArgGlyGlnAlaGlyPro-----GluSerL 1184
Db TCGTCTCGGGCTTCGGCGCGCGGTGCGGTACCTGTATGGGACATGCCACGAAAGACA 18612
Qy 1184 euGlyValProThrAlaGlnAlaSerValLeuArgGlyThrAlaLeuGlySerValP 1204
Db CCGGTGTCCGTACCGCGCAACGACTCGGGCGGATACCGGC---GTTCTCAGCGCCCTCC 18555
Qy 1204 roGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspAlaIle 1224
Db CAGGTGT-----CTCCAGGGGTGCGGTGCTGTCGGATC----- 18520
Qy 1224 hrTyrArgGlySerIleThrHisGlyThrProAla----- 1235
Db ---CATGCCAACCGCTCACCGCGGAGATCCGAAAGACCGCGGTGCAACTCGCCCGCC 18462
Qy 1236 -----AspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerPro 1253
Db CCGGCGATGAAGTTGCCAAGCGGTACGGGTGCGGTGCGGTGCGGAGACCACTCC 18402
Qy 1253 erArgLeuAsp-----ArgGly-----ArgGluAsp 1262
Db GCCGATCCAGCTCCCGGTCCCGGGTCCCGCGAGCGGTGTCGCCCGCGGAGACGATC 18342
Qy 1262 erLeuProLys-----GlyHisValIle 1270
Db AACTCCACAGCTGCTCCGGTGAATCCACTCCCGCGGACCGCGGAGGCGCATCCCGACG 18282
Qy 1270 yrGluGlyLys-----LysGlyHisValLeuSerTyrGluGlyMetSerVal 1287
Db ATGCGATCGGCTCGTTCGGGCGCACCGCTCGGGGCTTCGCCCGCGGCGCTCCCGC 18222

Qy 1287 hrGlnCysSerLysGluAspGlyArg-----SerSerSerGlyProProHisGluThrA 1305
Db CCGATCCCGCGCCCGCAGTTCCGGCGGCAGCGGTGTCGGCAGGCTCGTACGTCGGGTGG 18162
Qy 1305 laAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer--- 1323
Db TCGAAGACCGTCTGTCGTGGCCAGCCCGGTGTCAGTGTGCGTTCGCGAGTTCGCGAGT 18102
Qy 1324 -----SerAlaSerIleGluGly---LeuM 1331
Db TCAGGGCCGACAGCAGTCGACCGCGAGTTCGGCGNAGGCCCTGTCGACGGGACCTGC 18042
Qy 1331 etGlyArgAlaIleProProGlu-----ArgHisSerProHisHisLeuLys----- 1346
Db TCGGCGAGCGGTGGCGGACGCGCGTGTGTCGTACCGAGTTCGAGACCGCC 17982
Qy 1347 -----GluGlnHisIleArgGlySerIleThrG 1357
Db TTCTCGGTTCGCCGACGGGACCGCGGCGCCACCGCCCGCGCCCGCGGTTCGGGA 17922
Qy 1357 lnglyIle---ProArgSerTyrValGluAlaGlnGluAsp-----TyrLeuArgA 1373
Db CCGGGCGCGCCGACGCGGCTCTCGAGGGTGTCAAGAGCGGGTGGGGCGCTGGCG 17862
Qy 1373 rgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspL 1393
Db GT-----GTACGCGAGGAGAACCGTCCACCTGATGTGACGACGAT- 17818
Qy 1393 euThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHis----- 1411
Db CCGGGCTTACCTCCCTCCACCGACCGACCGTGCACCGCG 17784
Qy 1412 ---GluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlu---I 1428
Db AGACCTCGACGCGCTGTCT-----GGGTGTCATCTCATGACCCCG 17742
Qy 1428 leProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysG 1448
Db TGCGCGCGGAA-----CCGGTCCGCGACCGACCGCTCGGCC 17706
Qy 1448 lucGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrClys 1468
Db ATCCCGCTGCCCGCCAGGTACCCCA-----CGCCACCGAGGTG 17667
Qy 1468 erLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProValH 1488
Db GCGGGAGTCCCTCGCTGCTCTGTG-----GGCAGACCGTCGAGGTAGCG 17616
Qy 1488 isProLeuAspValMetAlaAspAlaArgAlaLeuGluArg-AlaCysTyrGluGluSer 1507
Db TTGCCCGGACGTAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG- 17559
Qy 1508 LeuLysSerArgProGlyThrAlaSerSerSerGlySerIleAlaArg---GlyAla 1526
Db ---AAGACACGAAACCGTCGAGGTCCCGGTCCCGGTCTGTCGAGGTTCGCGGCA 17502
Qy 1527 ProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGlu----- 1544
Db CCGACACCTTCGCCCGGTGGCCCGTTCGATCGGTCCCGGTGAGGTCTCGACCGGTG 17442
Qy 1545 -----AspHisGly----- 1547
Db CGTCTCGAGCGTTCGGCGCGAGTTCGAAACACCGCGCACAGCGGTGCTCGTCGACG 17382
Qy 1548 -----AlaProPheAlaGlyHisLeuProArgGlySerProValThrMet--- 1562
Db GTCCGAGGAGGCGACCGAGCTCTCCCGTTCGGTTCGTCGCGCGCGGTGAGCGGTGCC 17322
Qy 1563 ArgGluProThrProArgLeuGlnGlySerLeuSerSerSerSerSerSerSerSerSer 1582
Db CCGGTGCCAGGTTCGGCGAGTTCGGTGTAGTGTCTTCGACCCCGCTCGGTCC 17262

[illegible]

FT CDS 6592..7197
FT /tag= g
FT /gene= "megDIV"
FT /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT /note= "eryBVIII, dnmU homologue, encodes AAB82207"
FT 7220..8206
FT /tag= h
FT /gene= "megDV"
FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT /note= "eryVIV, dnmV homologue, encodes AAB82208"
FT complement(8228..9220)
FT /tag= i
FT /gene= "megBII-1(megDVII)"
FT /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT /note= "encodes AAB82209"
FT complement(9226..10479)
FT /tag= j
FT /gene= "megBV"
FT /product= "TDP-mycarose glycosyltransferase"
FT /note= "encodes AAB82210"
FT complement(10483..11424)
FT /tag= k
FT /gene= "megBIV"
FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT /note= "encodes AAB82211"
FT 12181..22821
FT /tag= l
FT /gene= "megA"
FT /product= "megalomicin 6-deoxyerthronolide B synthase 1"
FT /note= "polyketide synthase, encodes AAB82212"
FT 12505..13470
FT /tag= m
FT /gene= "megA"
FT /function= "AT-L"
FT 13576..13791
FT /tag= n
FT /gene= "megA"
FT /function= "ACP-L"
FT 13849..15126
FT /tag= o
FT /gene= "megA"
FT /function= "KS1"
FT 15427..16476
FT /tag= p
FT /gene= "megA"
FT /function= "AT1"
FT 17155..17694
FT /tag= q
FT /gene= "megA"
FT /function= "XR1"
FT 17947..18207
FT /tag= r
FT /gene= "megA"
FT /function= "ACP1"
FT 18268..19548
FT /tag= s
FT /gene= "megA"
FT /function= "KS2"
FT 19876..20910
FT /tag= t
FT /gene= "megA"
FT /function= "AT2"
FT 21517..22053
FT /tag= u
FT /gene= "megA"
FT /function= "XR2"
FT 22318..22575
FT /tag= v
FT /gene= "megA"
FT /function= "ACP2"
FT 22867..33555
FT /tag= w
FT /gene= "megAII"

FT misc_feature
FT /product= "megalomicin 6-deoxyerythronolide B synthase 2"
FT /note= "polyketide synthase, encodes AAB82213"
FT 22957..24237
FT /tag= x
FT /gene= "megAII"
FT /function= "KS3"
FT 22957..24237
FT /tag= ai
FT /gene= "megAIII"
FT /function= "KS5"
FT 24544..25581
FT /tag= aj
FT /gene= "megAIII"
FT /function= "AT5"
FT 26230..26733
FT /tag= z
FT /gene= "megAII"
FT /function= "KR3 (inactive)"
FT 26230..26733
FT /tag= ak
FT /gene= "megAIII"
FT /function= "KR5"
FT 26998..27258
FT /tag= aa
FT /gene= "megAII"
FT /function= "ACP3"
FT 26998..27258
FT /tag= al
FT /gene= "megAIII"
FT /function= "ACP5"
FT 27193..28590
FT /tag= ab
FT /gene= "megAII"
FT /function= "KS4"
FT 27193..28590
FT /tag= am
FT /gene= "megAIII"
FT /function= "KS6"
FT 28897..29931
FT /tag= ac
FT /gene= "megAII"
FT /function= "AT4"
FT 28897..29931
FT /tag= an
FT /gene= "megAIII"
FT /function= "AT6"
FT 29953..30477
FT /tag= ad
FT /gene= "megAII"
FT /function= "DH4"
FT 29953..30477
FT /tag= ao
FT /gene= "megAIII"
FT /function= "KR6"
FT 31396..32244
FT /tag= ae
FT /gene= "megAII"
FT /function= "ER4"
FT 31396..32244
FT /tag= ap
FT /gene= "megAIII"
FT /function= "ACP6"
FT 32257..32799
FT /tag= af
FT /gene= "megAII"
FT /function= "KR4"
FT 32257..32799
FT /tag= aq
FT /gene= "megAIII"

```
Db 54377 CATCACCCCGGTCACCGCCAGTCGCGGACGTCGCGATCAGGAAGTCTCGGC 54318
Qy 2134 -----ThrLeuAlaGlnHisIleSerGluValIle 2143
Db 54317 CACCTGCTCGCGGGGAACCGGATCGCGGACGTCGCCGACGAGGTGCGCGCGTC 54258
Qy 2144 ThrGlnAspTyr-----ThrArgHisHisPro-----GlnGlnLeu 2155
Db 54257 GCACCCGACACACGAGTCGCGCCGACCGCGCTCACCGCGCGCTGCGCGAGACGC- 54199
Qy 2156 SerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAsp 2175
Db 54198 CGGTACCCCGGAGGCTCTCGCGGAGT---CCAGCAGTCCACGCCCACTCCCGCG 54142
Qy 2176 LeuArg-----ArgProProSerAspLeuTyrLeuProProAspHisGlyAlaPro 2193
Db 54141 CGCGCCGACGCGCGGCGGCTCCGCGAGCGGCTCCACCTCCACCTGGGACACCA 54082
Qy 2194 AlaArgGlySerProHisSerGluGlyGlyIleGluProValSerProPro----- 2213
Db 54081 GCA-----GCGGCGCGCGCGCGGCTCCAGCGGCTCCACCTCCACCTGGGACACCA 54047
Qy 2214 ValLeuGlyGlyGlyGluAspGlyIleGluProValSerProPro----- 2228
Db 54046 GCAATC-----GGCATCACGACACGAGCTCGCTCGCTGTCATCCGC 54002
Qy 2229 -----GluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu 2245
Db 54001 ACCTCGCGCAGCGGATCACCAGCGCGGCGAGTCCGCCACGCG----- 53957
Qy 2246 TyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsn--- 2264
Db 53956 -----CCGACGCGCAGCAGCAGCCTCCAGCGCGCGCGGCTGC 53921
Qy 2265 -----ThrSerGlnProAlaPhePheSerLysLeuThrGlu 2277
Db 53920 ACGCCAGCGCGCGAGTGTGCGGGACGCGCGCGCGCTCGACGACCGCGGCTCC 53861
Qy 2278 SerAsnSerAlaMetValLysSerLysGlnGluLeuAsnLysLysLeuAsnThrHis 2297
Db 53860 ACCCGCAGCAGCGGCAAC-CCGCGCAGCGCAGCGCGCGCTCGCGCGCGCGCAC 53802
Qy 2298 AsnArgAsnGluPro---GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMet 2316
Db 53801 CAGCACCGGACCCCTCGACACCCCTCCGCTCCGCTCACCGCGCGCGCTCCCGCGAGC 53742
Qy 2317 ProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHieAla 2336
Db 53741 CGCGCCAGCTCGCGCACGTCGAGTCCGGGTACACCGGCGAGCGGTGACCGGCTCGAAC 53682
Qy 2337 SerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGln 2356
Db 53681 GCGGAGAACGTCGACACCGCGGGTTCCGCGCGAGCGGTTCCGCCACCGGTTGCGCGCG 53622
Qy 2357 TrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSer 2376
Db 53621 CTCAGAAAGTCCCGCGCAGCGCGCGACCGACCGACCTACATCTCCATGTCGCTCGGAGAGG 53562
Qy 2377 LeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThr----- 2393
Db 53561 TTCCCCGCTCGAGTGGCGCTCGCGCCAGCGCGGCGAGCGGCTCGCTCGTAGAAGGAG 53502
Qy 2394 -----LeuThrSer 2396
Db 53501 AAGATCAGCGCGCAGCTAGTGTGTCGCGGTCTTGAAGAAGTTCGATAGCGGTGAGC 53442
Qy 2397 ProGlyGlyGlyGlyAlaLysValSerGlyArgProSerSerArgLysAlaLysSer 2416
Db 53441 GCCTGCTCTCTCGCGGCGAGTTCGAGCGCGCGCGCGCGCTCGCGCGCTCCCGCG 53382
Qy 2417 ProAla-Pro-----GlyLeuAlaSerGlyAspAr 2426
```

```
Db 53381 CCGGTGGCCATGCCACACGACCCCGCGGAGAACATGGGTGCGCCGAGCAGCGCGCG 53322
Qy 2426 qPro---ProSerValSerSerValHis-----SerGluGlyVAspCysAsnArg-- 2441
Db 53321 TCCCGACCATCACCGACCGCGGCGGTGACTGCTGGTGGAGTGGTAGCTGTAGTGGTC 53262
Qy 2442 -----ArgThrProLeuThrAsnArgVal-----Trp----- 2450
Db 53261 TCCACCCAGAGTCCGAGGTGCGGAGTGCCTGAGCGGTGAGTGGTGTGATCTCTGGCACC 53202
Qy 2451 -----GluAspArgProSer-----SerAlaGlySerThrPr 2461
Db 53201 CGCTGACGTGCTCGCGAACCGCTCGCGCGGGTGCCTCGCGAGCGCGGTGCGCGGTGC 53142
Qy 2461 oPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVal----- 2475
Db 53141 ATGACCGTCCGACG-CTGATCGGTCCCGCCGACGAGGAGTCCGCCAGATCCACCGTC 53083
Qy 2476 -----MetAlaSerProProProGlyLeuProAlaGlySerGlyPr 2490
Db 53082 CGAGTGGTCCGACCTGATGTCGCTCGTGGCGGGGTGTGCGCGCTCGTCCAG-CC 53024
Qy 2490 oLeuAla-----GlyProHisIleAla-----TrpAspGluGluProLysPro 2504
Db 53023 CGTGGCGGTGGTGAACACCGCCACCATGCGAGGTCTCTGATCATCCCGCGAGCCC 52966

RESULT 69
AAF30757/C
ID AAF30757 standard; DNA; 47981 BP.
XX
AC AAF30757;
XX
DT 21-JUN-2001 (first entry)
XX
DE Micromonospora megalomicea megalomicin biosynthetic gene cluster.
XX
KW Megalomicin; meg gene; polyketide synthase; antibiotic; motilide;
KW antiparasitic; ds.
XX
OS Micromonospora megalomicea subsp. nigra.
XX
FH Key Location/Qualifiers
CDS complement (1..144)
FT /*tag= a
FT /partial
FT /gene= "megBVI (megT) T" /product="TDP-4-keto-6-
FT deoxyglucose-"
FT /note= "encodes AAB82201"
FT /note= "encodes AAB82201"
FT 928..2061
FT /*tag= b
FT /gene= "megDVI"
FT /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
FT /note= "encodes AAB82202"
FT 2072..3382
FT /*tag= c
FT /gene= "megDI"
FT /product= "TDP-megosamine glycosyltransferase"
FT /note= "eryCIII homologue; encodes AAB802203"
FT 3462..4634
FT /*tag= d
FT /gene= "meg"
FT /product= "mycarose O-acyltransferase"
FT /note= "encodes AAB82204"
FT 4651..5775
FT /*tag= e
FT /gene= "megDII"
FT /product= "TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
FT /note= "eryCI, DnrJ homologue, encodes AAB82205"
FT 5822..6595
FT /*tag= f
FT /gene= "megDIII"
FT /product= "daunosaminyl-N,N-dimethyltransferase"
FT /note= "eryCVI homologue; encodes AAB82206"
FT
```


QY 1504 -----TyrGluGluSerLeuLysSerArgProG1 1513
Db 56235 TCCCGCCCTGCGTCCCGCCCGCGCGCTCCGCTCCGCTACCCAGCTCGGCTCG 56176
QY 1513 yThrAlaSerSer-----SerGlyGlySerIleAlaArgGlyAl 1526
Db 56175 AACAGGATCGTCGGTTCGCGCCCGCTCCGCGCACCGGAAACAGGCTCGCGCGCGCGC 56116
QY 1526 aProValIleValProGluLeuGlyLys----- 1535
Db 56115 CCGTCCGCTCCCGCTCCCGCACGAGCTGTACTCCGCGCACCCCGCACCGCTCGCGCGCC 56056
QY 1536 -----ProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHi 1553
Db 56055 AACACCGCCCGCGCTCCCGCGGACCCCTTCCACGAGTCCGCGCA-----AG 56005
QY 1553 sLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlyLys 1573
Db 56004 CAGACCCCGTCCCTCCAGGTACCGCGCCAGCGCT-----CCCGCG----- 55962
QY 1573 rLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAl 1593
Db 55961 -----CCTTCCACCGCTCGTGGGTGTGTCGAGCGGAACCGCTCCCGCGCGCGGAGGCC 55906
QY 1593 aLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHi 1613
Db 55905 CCAGCACCGCCCGCTACGCTCTCGACCCCGCTCTGTCAGACGCCCA----- 55854
QY 1613 sLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAs 1633
Db 55853 -----GCGCGGCC 55846
QY 1633 pProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProAr 1653
Db 55845 TCCGCCAGCAGCGCCACCGGGGCTCCCGTCCAGCC-----CG 55807
QY 1653 gHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrIleuIleArgGlyTy 1673
Db 55806 CGCGCAGCACCGCGCCACCCAGTCGAGCGCAGGTCCCGCGC-ACGACCCCGCGTC 55748
QY 1673 rProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGl 1693
Db 55747 GCACGAGGCTCGCGCCCGACCTCCCGCACCGCGCTCTTGAACCCACCGACGTCGCG 55688
QY 1693 nGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLe 1713
Db 55687 CAGGTCCACCTCGAACAACACACCCCTCGCGCCAGGCGAGCGGAAACCGCGCGGCTC 55628
QY 1713 uSerProArgGluSerSerLeuAlaLeuAsn-----TyrAlaAlaGlyProArg----- 1729
Db 55627 CAACCCGGCGCCAGCAGCACACCTGTCGCGCACCCCGCGCGCGCGC-CCGCGCCAGCA 55569
QY 1730 -----GlyIleIleAspLeuSerGlnValProHi 1739
Db 55568 CCGCGTCGAAGAACACCGCTCTTGACCATGATGATGCTCGCTGATGCTGCGCGAGCGCCG 55509
QY 1739 sLeu-----ProValLeuVal-----ProProThrProGlyTh 1750
Db 55508 GATCACCGCCACCCGCTCGG-GCAGCAGCGGGTTCCCTCCCGCGCGCTCCAGGAAC 55450
QY 1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThr-----AlaProGlnPr 1767
Db 55449 GCGCGCCAGCGGTCTCGAACAACCGGTTCGCGCGCGCGGAGCTCTCTGCGCGCGCGCC 55390
QY 1767 o-----PheSerSerArgHisSerSerPro-----Le 1777
Db 55389 CCGCGTTCGCCAGCAGCTCCAGCTCTCGTCCGGCATCTCCACCGCGCCCGCGGTGT 55330
QY 1777 uSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgG1 1797
Db 55329 GAGTCCCGCTGGGA-----GGCTCATCGCAGCGGCC 55300

QY 1797 uArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSe 1817
Db 55299 CTTCCCC-----GACAGGCTCAGCGCGACCCAGCACCGCGGCTCCGCGGTCCG- 55243
QY 1817 rThrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySe 1837
Db 55242 GCGCGCCCGCGCACCGCAGGTCCGCTC-----CGGCGC-ACCAGCAGCACCTCCGCGCA 55188
QY 1837 rSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAl 1857
Db 55187 CCGCGCCCGCACCGCGCCAGCCACACGAGGTCTCCAGCGCAGCGCCCGCGCCCGC 55128
QY 1857 aHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerVa 1877
Db 55127 TCCACGCTCGACCCCGCCACCGC----- 55103
QY 1877 lLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrVa 1897
Db 55102 -----TCCGCGCCCGCACCGCCCGCGCG----- 55076
QY 1897 lLeuArgSerThrSerThrSerProValArgProAlaAlaThrPheProAlaTh 1917
Db 55075 -----CGCAGCGCCCGCTGCGCGTCTCCGCG-----CCCCCGCTCGT 55035
QY 1917 rHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLe 1937
Db 55034 GCACGCGAGTCCGCGCACCGGTGCGCGCGCGCGCGCGAT----- 54992
QY 1937 uLeuProLysGluAlaProArgValAlaArg-----ProGluArg-ProArgAlaAspThrG 1956
Db 54991 -----CCCCCGCTG 54936
QY 1956 lHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGlu----- 1970
Db 54935 GAGGCGCTCCGCGCACGAGAACCTTCCGCGCGCGCGCGGTTCAGCAGCGCGCATCACAC 54876
QY 1971 -----ProLaserSerProSerLysGlySerGluProArgPro----- 1983
Db 54875 CCGGTCCCGCACCGCGCACCGCGGTGTCTCCGCGCGCGCGCGCGCGCGCGCGCGCG 54816
QY 1984 -----LeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaL 2000
Db 54815 GTTCCCGCAGAGCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGTTCAGCA- 54757
QY 2000 yAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerA 2020
Db 54756 -----GCTCTCGGT 54705
QY 2020 spProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgS 2040
Db 54704 CTCCGCTCGCGCGCGCGCGGTTCATGCGCT-----GCCCGCC 54666
QY 2040 erLeuGlyTyrHisGlySerSerTyrSerProGluGlyVal-GluProValSerProVal 2059
Db 54665 GAACGGGT---GGTGTATCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGGTG 54609
QY 2060 -----SerSerProSerLeuThrHisAspLysGlyLeuProLysHis----- 2073
Db 54608 CGCTCGCGCGAGCTCGGTGGATCGGAACGACGACGACGACCTCCACCGCGCGGTGTCG 54549
QY 2074 -----LeuGluGluLeuAspLysSerHisLeuGlu 2083
Db 54548 CTCCGCGGGTGGCGGTGTCAGCGCGCGAGCTGGCGCGACCGACGCTCCCGCGCGCTGTC 54489
QY 2084 GlyGluLeuArgProLysGlnProGly-----ProValLysLeu 2096
Db 54488 GTCGCGCAGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 54429
QY 2097 GlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluGlnGlnProSerSer 2116
Db 54428 GCCGGAAGCGCGCACCGCTGACCATGTCCGCGCC-----GGCGCTCCAGCTGCC 54378
QY 2117 -----SerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValVal--- 2133

| | | | | |
|----|--|-------|--|-------|
| Qy | | 681 | ArgLysLeuGlyIysAlaProAlaAlaAlaSerGluGluAlaAlaPheProValVal | 700 |
| Db | | 59217 | CGGGTCCGGGTACGCCCGGCAGCACAGCAGCCGCGCTGACCGTCCACTCCGA | 59158 |
| Qy | | 701 | GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu | 720 |
| Db | | 59157 | CGGACGACGACCCGCTGGACCGG | 59131 |
| Qy | | 721 | AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla | 740 |
| Db | | 59130 | -----CACGCGACGGCGTGCTGCGCCCGCGGGCCACC CGAA----- | 59092 |
| Qy | | 741 | ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys | 760 |
| Db | | 59091 | -----CGCTACCGCGACGACTCCGAGCGCGCTC | 59065 |
| Qy | | 761 | AspThrGlyGlnAsnGlyProLysProAlaThrLeuGlyAlaAsp- | 776 |
| Db | | 59064 | CGGGTCGGGACCAATGGCGCTGGCGCGCGCT-----GGGGCTGATGTCCTCCGGCGGCG | 59011 |
| Qy | | 777 | -----GlyProProPro----- | 780 |
| Db | | 59010 | GGAGCCAATGGCGCTCTTGTCCTCCACGGGGCGGCACCCCGGGCGGACGACCAACGGT | 58951 |
| Qy | | 781 | ----GlyProProThrProPro-----ArgArgThrSerArgAlaProIleGluProTh | 797 |
| Db | | 58950 | GCGGCTCGCGCTGTCGCCGCCCGCGAGACGGCGCGGGCGCTGAGCTCGTGTGCCCGAC | 58891 |
| Qy | | 797 | rProAlaSerGluAlaThrGlyAlaProThrPro----- | 808 |
| Db | | 58890 | GGCGCGCGCCGACGGCTTCGGCTCGACCCCGCGCTCTGGAGCGCTGCTGCCGCGAC | 58831 |
| Qy | | 809 | -ProAlaProProSerProSerAlaProProValProLys----- | 824 |
| Db | | 58830 | CCCCCGCTGGCGGACGCTCTCTGCAGCCACCGGGGCGACCGCGCTCGGGTGGCG | 58771 |
| Qy | | 825 | -----GluGluLysGluGluThrAlaAlaAlaProValGluGluGluGl | 842 |
| Db | | 58770 | GGGACCTCGCGCTCTCGCGCTCGACGGCGGGGCGAC-CCGCTCGTGTCTGGTGGGGG | 58712 |
| Qy | | 842 | uGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluProVa | 862 |
| Db | | 58711 | TGTCACCCGAGTGGCGGCTCCCCGGCGCGCTCAGCGGGGTCTGTTTCGAGTCTGT | 58652 |
| Qy | | 862 | lLysSerGluCySthrGluGluAlaGluGlyProAlaLysGlyLysAspAlaGlu-- | 881 |
| Db | | 58651 | CGAGTCCCTTCGCCGCCCGCCCCCGCGCGCTCCACCGCCCTGTCCGTGACCT | 58592 |
| Qy | | 882 | -----AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGlu-LysLys----- | 897 |
| Db | | 58591 | GTCCACCGCCCTGTTCGGGACCTTCGGGGGTTCACGGCGCGAGGCGCTCGACGTCT | 58532 |
| Qy | | 897 | IuGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGln----- | 913 |
| Db | | 58531 | CCAGCGCTTCGGCGGAGCGCGCCCGCTGTCTGTGCGCACCGACCGCCCGACGA | 58472 |
| Qy | | 914 | -----AspSerAspSerSerAlaThrCySserAlaAspGluValAspGluAlaGlu | 931 |
| Db | | 58471 | CCCGTGGAGTGGCGCGCTCGGGCTGTGTGCGTCCGGCAGGCGGAGAACCCGGCGG | 58412 |
| Qy | | 931 | lyGlyAspLysAnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAasp | 951 |
| Db | | 58411 | GATCTGCTCGTCTCAGGACGGCGACGGCCCGAACCGCGGCCACACCGCGGGAACC | 58352 |
| Qy | | 951 | ro----- | 958 |
| Db | | 58351 | CGAGGTGGTCTGCCGCGACGGGGCGCTCCGCGTGC CGGCTGTGCCCGCACCGCGCCG | 58292 |
| Qy | | 959 | -----LysProLeuAspLeuLysGlnLeuLysGlnArgAlaAla | 972 |
| Db | | 58291 | CCACGACGGGTGCGCGCCCTCCCCGACACCGTCTCTGATCAGCGCGCGCACCGGGCGCT | 58232 |
| Qy | | 972 | laileProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProT | 992 |

| | | | | | |
|----|------|--|------|--|------|
| Db | 2147 | -ACCGGACCGC- ::: ::: | 1186 | TGATCACACCGCCACACCCCGCGCTCGTATGACCAATATTCGACTTCACCGAAC | 1127 |
| Qy | 1824 | aProIleThrArgProGlyThrGluGlnSerSerGlySer- ::: ::: ::: | 2119 | uLeuGln- | 2134 |
| Db | 2109 | CACACTCAACACACCCCGCCACCGCGCGGAGCAACTCACCCTGGGAATGACCCACCAC | 1126 | CCAACAACAACGCGCCACCCACAGACCCCGCCATACGACCCCAACACCGCTCAG-CC | 1068 |
| Qy | 1838 | ---SerGlySerSer- ::: ::: | 2134 | rLeuAlaGlnHisIleSerGlu- ::: ::: ::: | 2141 |
| Db | 2049 | CGATCCGGCACCACACCCCGGACATCCAAACCGCGCAACGACACCATCACCGCCCA | 1067 | TCATTCGGATACCCCAACCGAGTCCCGCTCCATCCACCATCCACATCCGAC | 1008 |
| Qy | 1842 | -GlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSe ::: ::: ::: | 2142 | -----Va | 2142 |
| Db | 1989 | CGACACCGGCTGCACCATCAATCAACCGACCC- ::: ::: ::: | 2142 | lIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPr | 2162 |
| Qy | 1861 | rProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnTh | 947 | GCGCGCGTCAACACCATTCGACCCACCATCTCTGATTTCACGCGAGAACCCCGCACCGCC | 888 |
| Db | 1947 | ACCGCGCGCGCGCGACCC- | 2162 | oLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAs | 2182 |
| Qy | 1881 | rGlyMetIysGlyIleIleThrAlaValGluProSerLysPro-ThrValLeuArgSerT ::: ::: | 887 | AACA- | 852 |
| Db | 1926 | CTCCAACAATCCCAATCCACAAACCCCGCAACCGCCGACCATCTCCA-CCATCGA | 2182 | pLeuTyrLeuProProAspHisGlyAlaProAlaargGlySerProHis- | 2199 |
| Qy | 1901 | hrSerThrSerSerProValArgProAlaAlaThrPheProAlaThrHisCys- ::: ::: ::: | 851 | CGCTCAACACACACCCCGCACCTCCCGCCACCCCGTCCCATCCGCACACCCCGCA | 792 |
| Db | 1869 | CTCGCGGAACACCTCCGACTCTCGAGCGCGCGCACCATATA-CCACACCATCTGCCAAC | 2200 | -----SerGluGlyGlyLysArgSerProGluProAs ::: ::: ::: | 2210 |
| Qy | 1920 | -----ProLeuGlyGlyThrLeuAspGlyValT | 791 | AAGCGCGGACCGGCCATCACCGGACACCCACCTCGCGAGAAACTCCACAAACAC | 732 |
| Db | 1810 | CCTGACCGGAAACACAAACACCCCGGACCCACACGACGACCAACCGCGCCACCG | 2210 | nLysThrSerValLeuGlyGlyGluaspGlyIleGluProValSerProProGluGl | 2230 |
| Qy | 1929 | yrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProG | 731 | CCCGCGG- | 705 |
| Db | 1750 | AACCCACACACCGAAGCC- | 2230 | yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgaspGlyGl | 2250 |
| Qy | 1949 | luArgProArgAlaAspThrGly-HisAlaPheLeuAlaLysProPro- ::: ::: | 704 | CCCGCAACCGGCATACACATCACCCACCGCAACCCCT- | 654 |
| Db | 1700 | --AGACAGGACAAACCTCTCTCCACACCCCGGACGACACCGCGCGTGTCTCAACA | 2250 | uGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProAl | 2270 |
| Qy | 1966 | ArgSerGlyLeuGluProAlaSerPro- ::: ::: ::: | 653 | TGCACGCGCCACCAACGACGACGACGCGGTATCAACCGMAATCGACGACCGCTCAAC | 594 |
| Db | 1642 | CGACCGCGGACGACACCAACGACCAACCCATCCACCAACCAACGACCAACCGCGAGC | 2270 | aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluI | 2290 |
| Qy | 1977 | LysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArg | 593 | C- | 593 |
| Db | 1582 | CACCAACCGCTCAGACACCCCGCGCACCT- | 2290 | eAsnLysLysLeuAsnThrHisAsnArgAsnGluPro- | 2309 |
| Qy | 1997 | ThrProAlaLysAsnLeuAlaProHisAlaSerProAspProProAlaProProAla | 592 | -----CCAACACATACGAACACGACCGGACACACACCTCGCGGACGTCCTCC | 546 |
| Db | 1534 | ACGCGCGGAAAGCC- | 2309 | oGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSe | 2329 |
| Qy | 2017 | SerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeu | 545 | GTCAACA- | 530 |
| Db | 1480 | AGCTCTGCAACACACCCCTCGCGCCA-TCCACGACACCGGCATCGGAATCGAATCG | 2329 | rGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAl | 2349 |
| Qy | 2037 | GluLeuArgSerLeuGlyTyrHisGlySer- | 529 | -----CAACCCACCCCGGACT- | 512 |
| Db | 1421 | GAATCGGAATCGCGAGATCGTCAACTCGGGGCTGCTCAAAATCAGATCGCATTC | 2349 | aLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAs | 2369 |
| Qy | 2049 | SerProGluGlyValGluProValSerProValSerProSerLeuThrHisAspLys | 511 | -----CCCAACCGCGCGCCCTACTCTCTGA | 486 |
| Db | 1361 | GTCCCGGACACCCCAACCGCGACACACCGC-ACGACGACACAG- | 2369 | nProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaaspGlyAr | 2389 |
| Qy | 2069 | GlyLeuProLysHisLeuGluGluLeuAsp- | 485 | CCCC- | 432 |
| Db | 1305 | CCAATCCCGCGCTCCGTCACCACTCCACCCCGGACCAACCAACCGGATCC | 2389 | gSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgPr | 2409 |
| Qy | 2080 | SerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlu | 431 | ATCCCGGACGCTCCACGCTCCCGACACCTCCAAACAAACAAACGCTGTCGGGATCC | 372 |
| Db | 1245 | CGGACATCATCATGCAACG-TGCGCGGACACACCCCGCGGACGACCATCATCT | 2409 | oSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSe | 2429 |
| Qy | 2100 | AlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerPro-Le | 371 | A- | 332 |

| | | | |
|----|------|---|------|
| Qy | 1247 | eGlyGluAspSerProSerArgLeuAspA:ggLyArgGluAsp----- | 1261 |
| Db | 3997 | GCCCGCGCGGCCAGGTCTGGGAGACCGAGGAAGACGACGAACGCTGCAGTCCA | 3938 |
| Qy | 1261 | ----- | 1261 |
| Db | 3937 | GTTCGGCCTCCGCGCAAGGGCGTGAGGTGCATCTCGTCGTCGTCTTCGCGCGCACGG | 3878 |
| Qy | 1262 | -----SerLeuProLysGlyHisValIleTyrgLuGlyLysLy | 1274 |
| Db | 3877 | CGCGCAGAAGTCTCTCCGGCGTGTCTCGCGAGGGCGGTGAGCGGACCGCGGCGGAG | 3818 |
| Qy | 1274 | sGlyHisValLeuSerTyrgLuGlyMetSerValThrGlnCyserLys-----Gl | 1292 |
| Db | 3817 | CCACCACA-TCGTCTGTGGCGCGTACGCGCCGAGCACCGCGCGAGCGGTCCCGETCG | 3759 |
| Qy | 1292 | wApGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArg-ThrTyra | 1312 |
| Db | 3758 | CTCGGGTCGAGTCGATCAGGGCCACCC-----GGTCGGCCCCCGCGGTAGAGCCCG | 3708 |
| Qy | 1312 | spMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetG | 1332 |
| Db | 3707 | AC-----GGTGGACGCGCGCGCGGCC----- | 3685 |
| Qy | 1332 | lyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisIleA | 1352 |
| Db | 3684 | -----GGCGAGCACCATCGTCG | 3669 |
| Qy | 1352 | rgGlySerIleThrGln-----GlyIleProA | 1361 |
| Db | 3668 | GCCCGCCGTCGCGGAGGAGCGCGCCGAGCGCGGAGCGCGGCGCGGTCTCCCC- | 3610 |
| Qy | 1361 | rgSerTyrgValGluAlaGlnGluAspTyrgLeuArgArgGluAlaLysLeuLysArgG | 1381 |
| Db | 3609 | -----GACGAC-----GAGG | 3600 |
| Qy | 1381 | luGlyThrProProProProProSerArgasp-----LeuthrGluAlat | 1397 |
| Db | 3599 | ATGTTGCCCGCGCGCCAGTCGCCGAGCGTGACGCTGGCGGGGAACACGACGAGCGGT | 3540 |
| Qy | 1397 | yrlYsThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlat | 1417 |
| Db | 3539 | CGACCTAGACCGCGGAGGCCG-----TACGGCGACTGGTCTCTCG | 3498 |
| Qy | 1417 | hrVallysGluAlaGlyArgSerIleHisGluIleProArGluGluLeuAurHisThrp | 1437 |
| Db | 3497 | CCGTT-----CGAGCGCCACACATCGG-CGAGCTGCGGAAAGCACGCTC | 3451 |
| Qy | 1437 | roGlu---LeuProLeuAlaPro-----ArgProLeuLysGluGlySerIleThrGlnG | 1454 |
| Db | 3450 | CCCGGTGAGCGCGCGAGTCCACACCGCCGCCCAT-CGGTCCGCGAGTTCGAGGGCCG | 3392 |
| Qy | 1454 | lyThr---ProLeuLysTyAspThrGlyAlaSerThrGlySerLysLysHisAspv | 1473 |
| Db | 3391 | CGACCCGCGGAGCCCCACACCTGGSGCTCCCGCGGTCCGCGACCTCCCGGAGGACA | 3332 |
| Qy | 1473 | al-ArgSerLeuIleGlySerProGlyArgThrPhePro-----ProValHisProLeu | 1490 |
| Db | 3331 | CGGCGACGGCCCCCGGTGACCGCCACGAGGCGGGCTCGCTCCCGCGTCCGCCAATG | 3272 |
| Qy | 1491 | AspValMetAlaAspAlaArqAlaLeuGluAargAlaCysTyrgLuGluSerLeuLysSer | 1510 |
| Db | 3271 | CCTGGATCAGGCGAGCGAGCGGTCAAGCGCGCGGTACGCGCGGGTGGCGCGTCCG | 3212 |
| Qy | 1511 | ArgProGlyThrAlaSerSer----- | 1517 |
| Db | 3211 | CGCCCTCGCGAGCGCCAGCAGGGAGACACGCCCGCGCGCGCCCTCGGGTCCG | 3152 |
| Qy | 1518 | -----SerGlyGly----- | 1520 |
| Db | 3151 | CGGGGCGCGGAGTTCAGGAGCGCGGAAGCGCGCCGCGCGGTGCCGACGGTCA | 3092 |

| | | | |
|----|------|---|------|
| Qy | 1521 | -----SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLys | 1539 |
| Db | 3091 | GCAGTTCGCGAGGCGCCACGCGCGCCAGCGCTCGCGTACGCCGCCA-----CCA | 3038 |
| Qy | 1536 | ProArgInSerProLeuThrTyrGluAspHisGlyAla----- | 1548 |
| Db | 3037 | CCCCGTCAACGGCCAGCGCTCGGGCCACCATCAGCAGCCAGCGTTGGTCTGAGGGGTCCGG | 2978 |
| Qy | 1549 | -----ProPheAlaGlyHisLeuProArg----- | 1556 |
| Db | 2977 | GGTCGTGGTCCGCAACGGTTTCCAGGTGATCCGATACCGCAGTCAACCCGCGCGGTCCG | 2918 |
| Qy | 1556 | ----- | 1556 |
| Db | 2917 | CGGTGTGTGTGCGCCGTGCGGGGTTGTGTGCGGGTTGTGTGCGGGTGTGTGCGCG | 2858 |
| Qy | 1557 | -----GlySerProValThrMetArgGlu | 1564 |
| Db | 2857 | GGTTGGTGGAGTCGGTCCCAAGTAGTGGGGGGTTTGGAAACGGATACGTGGCAGGTCAA | 2798 |
| Qy | 1565 | ProThrProArgLeuGlnGluGlySerLeuSerSerSertLysAlaSerGlnAspArgLys | 1584 |
| Db | 2797 | CCAAACGGCGCGCGCCCGCGGNACACACCTCTCCAGTCAACACTCATCCGCCGACGA | 2738 |
| Qy | 1585 | LeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHis | 1604 |
| Db | 2737 | ACGCCTCACCCACGACGTCA-----GAAACCGCGCCACCCACCAT | 2696 |
| Qy | 1605 | ProHisProIleSer-ProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTy | 1624 |
| Db | 2695 | CATCACGGCGCAGCAGCCACCCACCAACAC----- | 2668 |
| Qy | 1624 | rArgSerHisIleProLeuAlaPhe-----AspProThr-----Se | 1636 |
| Db | 2667 | -CGCGCCGACATCCGATCCACACCGCCACCGACTCCAAACCCCGCCGCAACACCG | 2609 |
| Qy | 1636 | rIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAl | 1656 |
| Db | 2608 | GATCGCCGGAACACTC-----CACAAACCCCGGAAAAACCCAT | 2573 |
| Qy | 1656 | aProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspTh | 1676 |
| Db | 2572 | CCCCAACAAAG-----CGCGCCCGCCACCATCCGAAACCAACACCCCGCT | 2531 |
| Qy | 1676 | rAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHi | 1696 |
| Db | 2530 | CACCGAGTTCGGATACCATACCCACATCCAAACCCACCCCATCAACACACCCACCA | 2471 |
| Qy | 1696 | sHis-----AsnThrAlaThrAlaMetAlaGlnAr | 1706 |
| Db | 2470 | CCACCGACGAATAAACCGCACACACCCCGCCCGGACACAAACCCCAACACCCCG | 2411 |
| Qy | 1706 | gAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsn---TyrAl | 1725 |
| Db | 2410 | ACAGCATCCCTCCACCGCGCTCCACATCGCGCGAATCGCAGCATATCCACCGCCACCC | 2351 |
| Qy | 1725 | aAlaGlyProArgGlyIle---IleAspLeuSerGlnValProHisLeuProValLeuVa | 1744 |
| Db | 2350 | GCGGACCCACACCCCATCCCGTTCACATCCAAACCCCAACAGCAGCAGCATCCA-CA | 2292 |
| Qy | 1744 | lProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAl | 1764 |
| Db | 2291 | TCCCCCGACACCA-----CGTCAACGACG-----GACCATTC | 2259 |
| Qy | 1764 | aProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHi | 1784 |
| Db | 2258 | ACCGCGCCACACCAACCGCACCAACCCACCCCGCCCAACCGTCTCTGCACATCCACGATC | 2199 |
| Qy | 1784 | sLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgArgGluAr | 1804 |
| Db | 2198 | GGCAGTCCGACCGACA-----CCATCGCACCAACCGGCCCAACCTCTACCGATC | 2148 |
| Qy | 1804 | qAspArgAspArgGluArgGlyLysSerIleLeuThrSerThrThrValGluHisAl | 1824 |

| | | | | | | | |
|----|------|--|------|----|------|---|------|
| Db | 6132 | CGCGCGCTCCCGCGCCGACAGGAACAGCGGACGGTCACTCGTCCGCTCGTGGTC | 6073 | Db | 5104 | GCAACGCGGTCCGGTCCAAACCGCGCGCTTCAGGGACTCCC | 5057 |
| Qy | 773 | euGlyAlaAspGlyProProGlyProPro | 783 | Qy | 1006 | GlnProGluSerAspAlaProGlnGlnPro | 1015 |
| Db | 6072 | GATCGTTCCGGACACACCGCGCGCTGCTCGACGATCACGTGGCGTGTGGTCCGCT | 6013 | Db | 5056 | TCTCCAGACGACGCGCTCTCGGATCATGGCGAGGCGCTCGCGGGGAGATCCCGA | 4997 |
| Qy | 784 | -----ThrProProArgArgThrSerArgAlaProIleGluProThrProA | 799 | Qy | 1016 | -----GlySerSerProArgGly | 1021 |
| Db | 6012 | GATCCCGAAGAGGACACCGCGCGCGGTTCGTGCGCCCTGCGGCCACGCGCACCG | 5953 | Db | 4996 | AGAACCGGGCATCAAGTCGGGCGCTCGTGAGGAAGCGCCCTCGCGACGTAGTGC | 4937 |
| Qy | 799 | laSerGluAlaThrGlyAlaProProPro-----ProAlaProProSerProA | 817 | Qy | 1021 | -----LysSerArg | 1024 |
| Db | 5952 | CTCGTTCAGCAACG-----ACACGCGCCCGAAGCCAGTCCACCCCGCGCGTCCG | 5902 | Db | 4936 | GGCGCGCGCTCCGGTCCGGTTCGTACAGCGCGTTCGACGTCCAGCCCCCGTTCGTGG | 4877 |
| Qy | 817 | laProPro-----ProValV | 822 | Qy | 1022 | -----LysSerArg | 1024 |
| Db | 5901 | CCGTTCACGTGCAGCGTCCGCGGACGGTGTCTGCGCGACGCCAGGATCACCTTCAT | 5842 | Db | 4876 | GGAACTGGAGATGGCGTCCGCGCCCTTCGCCAGCAGATCCACAGGTCTCGGGGCGT | 4817 |
| Qy | 822 | alProLysGlu-----GluLysGluGlu | 830 | Qy | 1025 | SerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGly | 1044 |
| Db | 5841 | CACCCCGCGATCCGCGCGCGCTGGTGTGACCGATGTTGACTTCAGCGAGCCGAG | 5782 | Db | 4816 | CGACCCCGCGGAAACCGACAGCCATCCGATCACCGCGATCGGTCTCGTGGCCG | 4757 |
| Qy | 830 | luThrAlaAlaAlaProProValGluGluGlyGlu----- | 841 | Qy | 1045 | AspProProCysTrpThrSerGly-----LeuProPheProValProProArgGluValle | 1063 |
| Db | 5781 | CCACAGTGGCGCGCGCGCGCGTCCCGCCGTACGTGGCGAGCGCTCGCGCTC | 5722 | Db | 4756 | CCCGCGCGCGCGAGCGCGGCGACGCGCTCCGCGTCTCTCGTCCGCGCTCGCGCC | 4697 |
| Qy | 842 | -----GluGlnLysProPro-AlaAla | 848 | Qy | 1064 | LysAlaSerProHisAla-----ProAspPro----- | 1072 |
| Db | 5721 | GATGGGTCCCGCGCGGTCCGCTGCGCGCTCGCGCTCCAGCGCTCCACCTGTCGG | 5662 | Db | 4696 | CGACGAGCTCGGCAGCGAGTACGCGGCCAGCACCCGGGCGCTGGGATGTCGAAGACGA | 4637 |
| Qy | 849 | GluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGlu | 868 | Qy | 1082 | HiePro-----LeuProLeuGlyLysHisAspThrAlaArgProValLeuPro | 1097 |
| Db | 5661 | CCGACGCGCGCTCGCGCAGCGCTCGCGATGATCCCGTTCCTGGGA----- | 5614 | Db | 4576 | TCACCGAGTGAAGCCGAGTTCCC-----GGAAACCGCGCTCGGATCGCGCGCTCGCGCT | 4520 |
| Qy | 869 | GluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGlu | 888 | Qy | 1098 | -----ArgProProThrIleSerAsnProProProLeuIleSerSerAla | 1112 |
| Db | 5613 | GGCGCGCTCGCGCGCGTGGCGCGTGGTGGCGCGCTCGCGCTCGCTGGTTCAC | 5566 | Db | 4519 | CGTCTGCCCG | 4460 |
| Qy | 888 | ----- | 888 | Qy | 1113 | LysHisProSerValLeuGluArgGlnIleGlyAlaIle-SerGlnGlyMetSerVal | 1131 |
| Db | 5565 | GGCGGAGCCCTTGACACCGCGCGAGTCTCGTCCCGCGCGCTCGCGTCCGCGAGCCG | 5506 | Db | 4459 | GATCG | 4400 |
| Qy | 889 | -----GlyAlaLeuLysAlaGluLysLysGluGlyLysSerGly--- | 901 | Qy | 1132 | -----GlnLeuHisValPro-----TyrSerGluHisAl | 1141 |
| Db | 5505 | CTGCACCAACCATATGCGCGCGCGCTCGTGTAGCGCCACCCCGTCCGCGCGCGCG | 5446 | Db | 4399 | CGCGCGCGGTGTCTCCGCTCCCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCG | 4340 |
| Qy | 902 | -----ArgAlaThr | 904 | Qy | 1141 | AlaValaProValGlyProValThrMetGlyLeuProLeuProMetAsp | 1157 |
| Db | 5445 | CGCTTCGACCGCGCTCGCGGACATCGCGCGTCCCGGAGAACTCCACGACATGTC | 5386 | Db | 4339 | TGGTGGCGCGCGCGCGGTGTACGAGCGCAGAACCGCGCGCGCGCGCGCGCGCG | 4280 |
| Qy | 905 | ThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAlaAsp | 924 | Qy | 1158 | ProLysLysLeuAlaProPheSerGlyValLysGlnGlu---GlnLeuSerPro---- | 1174 |
| Db | 5385 | CGCGTTCGACATCACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 5326 | Db | 4279 | TCACCGACCGCTCG | 4220 |
| Qy | 925 | -----GluValAspGluAlaGluGlyLysAspLysAsnArgLeuLeuSerPro | 940 | Qy | 1175 | -----ArgGlyGlnAlaGlyProProGlu-----SerLeuGlyValProThrAlaGlnG | 1191 |
| Db | 5325 | CGCTTCGACCGGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 5266 | Db | 4219 | CCACCGGTGAGCGCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 4160 |
| Qy | 941 | ArgProSerLeu---LeuThrProThrGlyAspProArg----- | 952 | Qy | 1191 | uAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG | 1211 |
| Db | 5265 | CGCGCGCGCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 5206 | Db | 4159 | CGGTTCGTCG | 4106 |
| Qy | 953 | -----AlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLysGlnArgAla | 970 | Qy | 1211 | yleProSerThrArgValProSerAspSerAlaIleThrThrArgGlySerIleThrHi | 1231 |
| Db | 5205 | GCGGTTCGCGAGTAGCTCGAAGTCTCGCGCGCGCTCGACGACGCGCGATGATG | 5146 | Db | 4105 | TCGCGTCCG | 4058 |
| Qy | 971 | AlaAlaProProIleGlnValThrLysValHisGluProProArgGluAsp----- | 988 | Qy | 1231 | sglyThrProAla-----AspValLeuThrLysGlyThrIleThrArgIle----- | 1247 |
| Db | 5145 | CTGGCGGTGTGTC-----CAGGAACA-CCCGGTTCGCGCGCGCGCGCGCG | 5105 | Qy | 4057 | GGCGCGCGCTCTCTCTCGACCGCGGTTCACCGCGCGGTTCGCGCGCGCGCGGT | 3998 |
| Qy | 989 | -----AlaAlaProThrLysProAlaProProProProProProProProProPro | 1005 | | | | |

Db 18091 --ACCAGGAGCTCGGCTCTACGCCCCACAGCCACACGCGGTCCAGCGCTACTCCA 18034
 Qy 2400 -----GlycyllysAlaLysValSerGlyArg--- 2408
 Db 18033 CCGGAGAGCGGGGTGGTGAACCGCTCTGTCAGCAACAACGACTCGCGCTCC 17974
 Qy 2409 -----ProSerSerArgLysAlaLysSerProAlaProGlyLeuAla----- 2422
 Db 17973 CCGCTCCCGCCACATCACCTCGCGCAGCGGGGTCCAGCTCCCGGTGGAACAGCGTCA 17914
 Qy 2423 ---SerGlyAspArgProProSerValSerVal-----HisSerGluGlyAspCys 2439
 Db 17913 CGACCGGTGAACCGCTCTCCGGAACGCTGCCACGCGCGCGCAGACCCCGGCCCATGC 17854
 Qy 2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
 Db 17853 CCGCGCTGTCGG-----CGCCCTGTCGGTGAACAGCA 17821
 Qy 2460 ThrProPheProTyrAsnProLeuLeuMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
 Db 17820 ACGCCA-----GCTTGGCGCGGAGAGCTCGCATGCGAGCGGCCCG 17779
 Qy 2480 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAsp 2499
 Db 17778 CCCC CGCGCGCTGTCCTCCGCGCCACGCGCGGCGGCGCGCAGCGCCCTCGCGCG 17719
 Qy 2500 GluGluProLysPro 2504
 Db 17718 ACGTCACCGCCACCG 17704

RESULT 67

ID AAD55817 standard; DNA; 11238 BP.

AC AAD55817;

XX 27-OCT-2003 (revised)
 DT 07-AUG-2003 (first entry)

XX Micromonospora carbonacea polyketide synthase (PKS) type I gene #7.

XX Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
 gene; ds.

XX Micromonospora carbonacea.

XX Key Location/Qualifiers
 FH 1..11238
 CDS /tag= a

FT /product= "polyketide synthase"
 FT /note= "CDS does not include start codon"
 FT /partial

XX CA2391131-A1.

XX 19-NOV-2002.

XX 26-JUL-2002; 2002CA-02391131.

XX 26-JUL-2001; 2001US-0307629P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Yang X, Staffa A, Farnet CM;

XX WPI; 2003-343556/33.

XX P-PSDB; AAE37001.

XX Novel isolated polypeptide involved in biosynthesis of macrolides by
 PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
 PT preferably for biosynthesis of rosaramicin.

XX Claim 1; Page 151-158; 206pp; English.

XX The invention relates to genes and proteins involved in the biosynthesis
 CC of macrolides by microorganisms. In particular it relates to the nucleic
 CC acids forming the biosynthetic locus for rosaramicin (a 16-member
 CC macrolide antibiotic) from Micromonospora carbonacea. The invention is
 CC useful for the biosynthesis of macrolides by microorganisms. It allows
 CC direct manipulation of the proteins involved in the biosynthesis of
 CC chemical engineering of the proteins involved in the biosynthesis of
 CC rosaramicin. It is useful to catalyse certain biochemical reactions, in
 CC vitro or in vivo, to direct or enhance the synthesis or modification of a
 CC polyketide, polyketide substrate or its precursor. The present sequence
 CC is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-
 CC OCT-2003 to standardise OS field)

XX Sequence 11238 BP; 1110 A; 3596 C; 4707 G; 1825 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,42e-10 Length: 11238
 Score: 622.00 Matches: 656
 Percent Similarity: 29.24% Conservative: 296
 Best Local Similarity: 20.15% Mismatches: 1118
 Query Match: 4.71% Indels: 1197
 DB: 7 Gaps: 144

US-09-522-753-5 (1-2517) x AAD55817 (1-11238)

Qy 3 GlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHis 22
 Db 8568 GGTGGCGCGCACCGCGTCCGGCCCGACCGCGCGCGCGCGGACCGCGCACCGCGCAC 8509
 Qy 23 SerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyr 42
 Db 8508 -----GTCCGGCAGCGCGCGCC 8491
 Qy 43 GlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleLeuGlnProGln 62
 Db 8490 TTCGGCCACCGCGCGCACCGCGCGCGAGATCGGG-----GTACGCCCGCA 8443
 Qy 63 ArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeu 82
 Db 8442 GAGGCCCGCGCGGTCCAGGCCCGCGCGCGCGGATCGCGCGCGC----- 8392
 Qy 83 HisLeuArgProGluSerHisSer----- 90
 Db 8391 CACGACCGCGCAGCGCTGCACCGCGCGCACCGCGCGGTTCGCGCGCGCGGTCCAGCGCAC 8332
 Qy 91 -----TyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
 Db 8331 GTGGTGAGCGCGCGGTCCCGGTCCCGGGGTCCCGGGGTGAGCGCTCCACCGGCAC 8272
 Qy 107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThr 126
 Db 8271 CGCGCGCGCCACCGAGGAGTC---GATCACACGACCGGGTTGCCCGCTCGTCGACGCA 8215
 Qy 127 GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeu--- 145
 Db 8214 GCGCATCCG-----GACCGCTCTCGGCACCGCGCGCGCTGAT 8179
 Qy 146 -----GluProValSerProProHisThrAspProGluLeuGluLeu 162
 Db 8178 GGTACCGCGCAGCGGTCCGCCCGCGCACGACGACCGCGCGCGCGCGGCGGACGG 8119
 Qy 163 ValPro-----ProArgLeuSerLysGlu-----GluLeuIleGlnAsnMetAspArg 178
 Db 8118 CATCCGACCTGCGCGCTCTCGGGGAAGACGCGCGCGAGGCTCATCTCGAAGCGCGCG 8059
 Qy 179 ValaspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGln 198
 Db 8058 GTCGAGCAGCGCGCGGTGGACACCGACGCGCGCTCCCGGGCTCCGGGACGGGCGGCG 7999
 Qy 199 GlnLeu---GluGluGluAlaAlaLysProPro-----GluProGluLys--- 212
 Db 7998 CACCTCGCGGTACACCGTGTCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 7939

| | | | | | | | | |
|----|-------|--|--------------------------------------|-------|-------|--|---|-------|
| QY | 1814 | IleLeuThrSerThrThrThrValGluHisAlaProLeu | ----- | 1826 | Db | 19051 | CATCCCTGCGACGAAAGCCGCCGCGCTCCCGACTCCACCCCGGTGGCCAGC | 18992 |
| Db | 19975 | CGGGTAAGGTGTGCAGCACC | -----CATGCCCTCCACCTTGGGGCGCAACACCCGG | 19922 | QY | 2112 | erGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgV | 2132 |
| QY | 1827 | -----TTP | -----ArgProGly | 1830 | Db | 18991 | ACCAGCACCCCGCGACCG | 18959 |
| Db | 19921 | GCAGCGGACCGGCTCTGGTGGCCAGCAGCGCGTGTGAGCAGACCCCGGGGTGC | 19862 | QY | 1831 | -----ThrGlnSerSerGlySerSerGlySerGlyGlyGlyGlySerSer | 1849 | |
| QY | 1961 | -----ThrGlnSerSerGlySerSerGlySerGlyGlyGlyGlySerSer | 1849 | Db | 1961 | ACGACCCCGCGACCGCGCTCGACGGCGCCCAAGAGCGCCATGCTTCCGATCG | 19802 | |
| QY | 1850 | -----Arg-ProLaserHisSerHi | 1856 | Db | 19922 | CGGTAGAACCACTGCCAGCGCATCGGGCGGTGCTCCGAGCCCTCCGAGCGCTCGATCGCC | 18863 | |
| QY | 1856 | 8AlaHisGlnHisSerProLaser | -----ProArg | 1866 | QY | 2152 | ro | 2157 |
| Db | 19741 | GCAATGCGCGCGCACCTCTGGCGGTGTATCTCGGCCCCATTCTCGCGTGGGCAAT | 19682 | QY | 1867 | -----ThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnTh | 1881 | |
| QY | 19681 | CCGTGCGGTGTATCAGCACAGGTGCCCCCGCGCTCCACCAACACCGCGCAACG | 19622 | QY | 2176 | euArgProProSerAspLeuTyLeuProProAspHisGlyAlaProAlaArg | 2196 | |
| QY | 1881 | rGlyMetLysGlyLleThrAlaValGluProSerLysProThrValLeuArgSerTh | 1901 | Db | 19621 | AGGAGCCCAAGG | -----CGCCAGCCACCG | 19589 |
| QY | 1901 | rSerThrSerSerPro | -----ValArgProAlaAlaThrPheProProAla | 1916 | QY | 2229 | lueGlyMetThrGluProGlyHisArgSerAlaValTyProLeu | 2244 |
| Db | 19588 | ANGTAATCCCTCCGCGAGACGACCGCGGCTGCTTCCCTCGGTGGGGGGGCC | 19529 | QY | 18733 | CCGCGAGCTCTCCGCGCGCGCGCGCAAGC-ACGCAGCAACGTCGGTCCGCTCCGG | 18675 | |
| QY | 1917 | -----ThrHisCysProLeuGlyGlyThrLeuAspGlyValTyProTh | 1931 | QY | 2245 | -----LeuTyArgAspGlyGluGlnThrGluProSerArgM | 2257 | |
| Db | 19528 | ACGAGCGCGCTGGCGCGCGCCCTGGCGGATCCAGCTGAT | 19483 | QY | 18614 | CGCGCGGCTTGTGAGCGCGCTTACCCCGTCCGGAAGCGCACCGCTCCGCGACGTG | 18555 | |
| QY | 1931 | rLeuMetGluProValLeuLeuProLysGlyAlaProArgValAlaArgProGluArgPr | 1951 | QY | 2257 | etGlySerLysSerProGlyAsnThrSerGlnPro--ProAlaPheSerLysLeu- | 2275 | |
| Db | 19482 | -----CCTCGCGTCCGCGGCGAAGAGCTCGGCGACAGGGCC | 19445 | Db | 18554 | CCGACCCAGTAGCCCGCGCGCTAGCTCGTCCGTGACACCC-TCCTCCGCTCAGGTGC | 18496 | |
| QY | 1951 | oArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPr | 1971 | QY | 2276 | -----ThrGluSerAsnSerA | 2281 | |
| Db | 19444 | TCGACCTCGTGGGCTCTCTCCGATCCAGGTCCACAGCGCCCGCCAGAGCGCGGA | 19385 | Db | 18495 | TCACGACGAAACCTTGGCGCGGTACGTACACGACGCGCACCCGCGCAACTCTCT | 18436 | |
| QY | 1971 | oAlaSerSerSerLysGlySerGluProArgProLeuValProProValSerGlyHi | 1991 | QY | 2281 | laMetValLysSerLysGlnGlnLeuAsn | 2299 | |
| Db | 19384 | TGCTCCAGCGCGCGACCGGC | -----CCATACCCACAGCGCGCTGACAGGGGCA | 19331 | Db | 18435 | CCAGCATCGTTCCATCAGCGCGAGTGAACCGGTGCGAGACATGCGCGCTTGTGC | 18376 |
| QY | 1991 | 8AlaThrIleAlaArgThrProAlaLysAsnLeu | -----AlaProHisHi | 2006 | QY | 2299 | rgAsnGluProGluTyAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaI | 2319 |
| Db | 19330 | CGGTACGGCTCGCGCCACCGTGCAGCGCCCTCGGGTCCAGTCCAGAGCGGGGTGAG | 19271 | QY | 18375 | GCAGCGCGCGCGGACCCCGCGCGATCGCTGACCGCTGCTCCACCGCGCGA | 18316 | |
| QY | 2006 | 8AlaSerProAspProAlaProProAlaSerAlaSerAspProHisArgGluLysTh | 2026 | QY | 2319 | leThrGly--ThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHisAlaSerT | 2338 | |
| Db | 19270 | CGCGGCGCGCGCGCTGATCAGCGCGAGCACCGCGCGGTGGCAAGATGGTG | 19211 | Db | 18315 | TCACCACTCTCGCGCGCATTTGACC | -----GCCGCGATCGACACCGCGCGC- | 18266 |
| QY | 2026 | rGlnSerLysPro-PheSerIleGlnGluLeuArgSerLeuGlyTyHisGlyS | 2046 | QY | 2338 | hrAsnMetGlyLeuGluAlaIleAlaArgLysAlaLeuMetGlyLysTyAspGlnTrp- | 2357 | |
| Db | 19210 | ACTTTCGCGACTTTCGCG | -----CGATCGCCCGCTCCAGACG | 19169 | Db | 18265 | -----TCGCGCGCACCGCGCGCCACTCCCGCTCCGCGGTCCGAGGACACCATCG | 18211 |
| QY | 2046 | erSerTySerProGluGlyValGluProValSerProValSerSerProSerLeuThR | 2066 | QY | 2358 | -----GluGluSerPro | -----ProLeuSerAlaAsnAlaPheAsnPro- | 2370 |
| Db | 19168 | CGGTCCAGACCCACAGGTACAGCACCCCTG | -----CCAGTGTTCGG | 19124 | Db | 18210 | CGCGCGCGCGAGAGCCCTGTATCAGCGCGCGCGCGCGCGCATCACCCTAT | 18151 |
| QY | 2066 | isAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGluL | 2086 | QY | 2371 | -----LeuAsnAlaSerAlaSerLeuProAlaAlaMetProI | 2383 | |
| Db | 19123 | CCACCGAGGCGCTG | -----GTCACCTCTCGCGCAC- | 19092 | QY | 18150 | CTTCCAGCGAGAACACCCCGCGCACGCGCGCACGCTCCCGATGATGCCA- | 18092 |
| QY | 2086 | euArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHis | 2102 | QY | 2383 | leThrAlaAlaAspGlyArgSerAspHis-ThrLeuThrSerProGlyGly- | 2399 | |
| Db | 19091 | -----CGCGGAGC | -----CTCGCGGCGCATGAGCAGCGCGCGCACG | 19052 | QY | 2399 | ----- | 2399 |
| QY | 2103 | -----LeuProHisLeuArgProLeuGluS | 2112 | | | | | |

| | | | | | | | | | |
|----|-------|---|---|----|-------|--|---|-------------------|------|
| Db | 21701 | GCGGGCTTCCCGTACACGTCGCTCAGCGCCCGCACCTCGATCGGGTCGCCCGCAGCGCGCT | 21642 | Qy | 1472 | AspValArgSerLeuIleGlySerProGlyArgThrPhePro | -----Pro | 1486 | |
| Qy | 1184 | uGlyValPro | -----ThrAlaGlnGluAlaSe | Db | 20760 | GGTCTCTGACACCCCGCCGGAGC | -----GGCAGCGGCACCCACGATGCGGATCGGTCT | 20704 | |
| Db | 21641 | CCGGTCCCGTGGCATTCACGAAATCGAGTCGCGCGGAAACCACTCTGCTGCGCCAG | 21582 | Qy | 1487 | ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGlu | ----- | 1506 | |
| Qy | 1193 | rValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIle | ----- | Db | 20703 | CTGTGCGCG | -----GCAACCGAACGC | 20683 | |
| Db | 21581 | CGCTGCGGTAACAGCGCTCTCTGGGAGCGCGCTGGGCACTGTGAGCCCGCTGCTCGG | 21522 | Qy | 1507 | SerLeuLysSerArgProGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAla | ----- | 1526 | |
| Qy | 1208 | ----- | ----- | Db | 20682 | ---TTGCGACGTGCGGTCTCTGAGCTCCAGCTCCAGCAAGT | -----ATT | 20632 | |
| Db | 21521 | GCCATCATGATTGATCGCGTACCCCGGATCACCGCAGCATGGGTGCTGCGTTCGCGTG | 21462 | Qy | 1527 | ProValIleValProGluLeu | -----GlyLysProArgGlnSerPro | -----LeuThrTyrGlu | 1544 |
| Qy | 1209 | -----ThrLysGlyIleProSerThrArgVal | -----ProSe | Db | 20631 | CCACGACCGCTCCACGTCGGATGGTTCGAAACGCCAGCGTCGCCGACAGCGGCATACCCA | 20572 | | |
| Db | 21461 | CGCGTCGAGAGCGCTTGAGCACCAACCGCGCAGCGCTCGGCGCGCGTAGCCGCTC | 21402 | Qy | 1545 | AspHis | -----GlyAlaProPheAlaGlyHisLeuProArgGlySerPro | 1559 | |
| Qy | 1219 | rAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAsp | -----Valle | Db | 20571 | GCTCACCTGAAGCGCTTTCGGGATCTCCA | -----CAGCCATCAGGAGT | 20527 | |
| Db | 21401 | CGCGTCGCGC | -----GAGAACGCTTTCACACCGCCCGCGCGGCGGAAAGCGCGTG | Qy | 1560 | ValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerIysAla | ----- | 1579 | |
| Qy | 1238 | uTyrLysGlyThrIleThrArgIleIleGlyGluAspSer | -----ProSerAr | Db | 20526 | -----CGAGGCGCTCTCCACACAG | -----AGCTCTGCTCGACGTCGAGCG | 20488 | |
| Db | 21353 | CATCCGTGAGAGCAGCGGAGGCTTCGCGCAGCAGCATGTTCACCCCGCAACAG | 21294 | Qy | 1580 | SerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThr | 1599 | | |
| Qy | 1254 | q-LeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysL | ----- | Db | 20487 | TGCTTGTGTCGG | -----TAAAGCCCATCACCCCGCGCAGCACCGCGCACCAAGCTCGTAGA | 20431 | |
| Db | 21293 | GCTTGTGTCGACTCGCCCC | -----GTCCGAAGCTCTGCGACGC | Qy | 1600 | ValProGluHisHisProHisProLysSerProTyrGluHisLeuLeuArgGlyValSer | 1619 | | |
| Qy | 1274 | ysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSer | -----LysG | Db | 20430 | CGCGGAGCGGTCTCCACACAG | -----ACGCTTTCGCGCAGC | 20392 | |
| Db | 21254 | GAGGT | -----GCAGCGCACGAGCAGCAGCAGCGCGCTATCCAC | Qy | 1620 | GlyValAspLeuTyrArg | -----SerHisIleProLeuAlaPheAspProThrSerIlePro | 1638 | |
| Qy | 1292 | luAspGlyArgSerSerGlyProHisGluThrAlaAlaProLysArgThrTyrA | 1312 | Db | 20391 | GCTCTACAGCTGCGCAGCGCGAGGAGCGCTTTCGTGACAGTTACCGCGCATCCC | 20332 | | |
| Db | 21212 | AGCCAGGTGCGCGCTGACGCCCGCAAGAAATGACAGCGCTCCCGCGCAACGCTGAG | 21153 | Qy | 1639 | ArgGlyIleProLeuAspAlaAlaIleTyrTyrLeuProArgHisLeuAlaProAsn | 1658 | | |
| Qy | 1312 | spMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetG | 1332 | Db | 20331 | AGAAGC | -----GGCCTCGCTGCGTCCGCGGAGCGCTCCCGCAT | 20281 | |
| Db | 21152 | CATGTTCCCGTGCCTGT | ----- | Qy | 1659 | ProThrTyrProHisLeuTyrProTyrTyrLeuIleArgGlyTyrProAspThrAlaAla | 1678 | | |
| Qy | 1332 | lyrArgAlaIleProProGluArgHisSerPro-HisHisLeu | ----- | Db | 20280 | CCATCT | -----GGATGACACCGCGTGGCTCG | 20254 | |
| Db | 21132 | ---AGAGCCCCCGCGCTCATCGCGAGGTCTGACCCCGCTCGCATATTCTGTTGGGCGC | 21075 | Qy | 1679 | LeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsn | 1698 | | |
| Qy | 1346 | LysGluGlnHisHisIleArgGlySerIleThrGln | ----- | Db | 20253 | CGCGCGTACCGACGAGCC | -----ATTCCATCGCGCGCGCAGCGCAGCTCG | 20209 | |
| Db | 21074 | CGCGCCACGACACGCGCTGGGCTCTCGCGCAGCGCGCGGCTCTGCGCGCGCGCG | 21015 | Qy | 1699 | ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer | 1718 | | |
| Qy | 1358 | -----GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAla | 1375 | Db | 20208 | TCGGCA | -----TCGCGCCAGA | 20194 | |
| Db | 21014 | CTCGATCGCTCCAGCTCACTTC | -----CAGCAGCAGCGCGTGTTCGGGTCCAGGCT | Qy | 1719 | SerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValPro | 1738 | | |
| Qy | 1376 | LysLeuLeuLysArgGluGlyThrProProProProProProProProProProProPro | -----CCGAGCGAGATGTGGAAGA | Db | 20193 | ----- | -----TTCCCGCATGCT | 20182 | |
| Db | 20960 | CATCGCCTC | ----- | Qy | 1739 | HisLeuProValLeuValProThrProGlyThrProAlaThrAlaMetAspArgLeu | 1758 | | |
| Qy | 1396 | AlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGlyLeuVal | 1415 | Db | 20181 | CATGTTCCGCGCTGCGCTCGAGCCATCCCGCTCCCGCTCCG | ----- | 20140 | |
| Db | 20930 | CGCGCATCAAGCTCTCCACCTCGCGCAGAAAGCCACCCCTGGGCGAGTGTGTCT | 20871 | Qy | 1759 | AlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSer | 1778 | | |
| Qy | 1416 | AlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHis | 1435 | Db | 20139 | -----CCACAGCGCCCGAGGATGTCTCAGGCGCGCA | -----GCCCTTCGCTTC | 20095 | |
| Db | 20870 | CGGAGCCTCTCCGAGCGCGGCG | -----CGGCGCGCTCTGCCCAT | Qy | 1779 | ProGlyGlyPro | -----ThrHisLeuThrLysProThrThrThrSerSer | 1793 | |
| Qy | 1436 | ThrProGluLeuProLeu | -----AlaProArgProLeuLysGlySerIle | Db | 20094 | GACGAGGTCTCGCGCGCTGCCAAAGGCGATTCCTGCGCGTAGTGTGCTCCCTGGCGCA | 20035 | | |
| Db | 20829 | TCCACCGCTCGCGCGCGCTCGGTGCTGACACCAACCGCCCT | ----- | Qy | 1794 | SerGluArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSer | 1813 | | |
| Qy | 1452 | ThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrThrThrThrThr | ----- | Db | 20034 | TCGAGCCGAGACG | 19976 | | |
| Db | 20787 | -----CGGTCAACAGCTCCCGTAGTACTCCA | 20761 | | | | | | |

```
QY 583 hrArgSerMetAlaAsnGluAla-----AsnSerGluGlu- 594
DQ 23557 GCGCGGCATGGAGCAGCAGCATGGAAGTCCACGTGTGCACAGCGCTGCAGCGCCGC 23498
QY 595 AlaIleThrProGlnGlnSerAla-----GluLeuAlaSerMetGlu 608
DQ 23497 TCGCCGACTCCACCTTATCCGCCAATACACAGCCAGCTCCCGCGGTTCGCTCTCTGAT 23438
QY 609 Leu-----AsnGluSerSerArgTrpThrGluGluGluMetGlu 621
DQ 23437 TTCTGGAGGCTGCGAGTATCTCCGCCAGTCCACGCGATAGAACACAGCGCTTGGGTCCGA 23378
QY 622 ThrAlaLys-----LysGlyLeuLeuGlu----- 629
DQ 23377 TCGCCGCTGCTGCGGACGCGTGTCCAGGCTCATGTTCGGCGCAGTACCGCTCCCGCTGC 23318
QY 630 HisGlyArgAsnTrpSerAlaIleAlaAaGMetValGlySerLysThrValSerGlnCys 649
DQ 23317 CACGGATAGTCCGCGAGCGCACCCCGCGCCAGCCGCTGGGGAGAGACCCCGGCCAG--- 23261
QY 650 LysAsnPhetyrPheAsnTyrlsYsArgGlnAsnLeuAspGluIleLeuGlnGlnHis 669
DQ 23260 -----CTGACCGAGCGCGCGGCCAC 23237
QY 670 LysLeuLysMetGluLysGluAraGsnAlaAraArgLysLysLysAlaProAlaAla 689
DQ 23236 AGCCTG-----CCAGCGCC 23222
QY 690 AlaSerGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGly 709
DQ 23221 TCAGACACCCCGAGCGCTCTCCGCGCGCGCGCAACAGCAGCCAGCGCTCGCGTCC 23162
QY 710 ValSerGlyAsnGluGluGluMetValGluGluAlaAlaLeuHisAlaSerGlyAsn 729
DQ 23161 GCCTCCGCGAGGCAA-----GCTGGCAAC 23138
QY 730 GluValPro-ArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGly 749
DQ 23137 AGG---CCGAGCAGCGTCGCTTTCGGGCCCACTTCGAGGAAACGTCCTCCGCGCGCTTCG 23081
QY 749 uSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
DQ 23080 TGACGCGCTTACCCCGTCCGGAAGCGCACCGCTCCCGCACGCTCCCGCACCCAGTAG 23021
QY 769 oProAlaThrLeuGlyAlaAspGlyProPro----- 779
DQ 23020 CCGCGCGCTCAGCTCGTCCGTCGACCACTTCCCGCTCAGTTGCTCACCGAGAAACG 22961
QY 780 -----ProGlyProPro----- 783
DQ 22960 CTTGGCGCGGTACGTACCGACGCGCGCACCGCCCGCCGAACTCTCTCCAGCATCGGTTC 22901
QY 783 ----- 783
DQ 22900 ATCAGCGCGAGTGGGATGCGTGAGACATGACGCGCTTGGTGGCGACCGCGCGGCC 22841
QY 784 ----ThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAl 802
DQ 22840 GCGNACCCCGCGGA-----TCGCTGACCGCTTGTCTCCAGCGCGGATCACCAAC 22787
QY 802 aThrGlyAlaProThrProProAlaProProSerPro---SerAlaProProProVa 821
DQ 22786 TGCTCCGCGCCATTTGACCGCGCGATTCGACACACGCGCGTGGCGCGCACCGCGCG 22727
QY 821 lValProLysGluLysGluGluThrAlaAlaAlaProProValGluGluGlyGly 841
DQ 22726 GCCACCTCCGCGCTCCGAGCGACACC-ATCGCGCGCGCGCGCGAGAGCCCGCTG 22668
QY 841 uGluGlnLysProProAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluPr 861
DQ 22667 CATCAGCGCGCGCGCGCGCGCGCGCTCACC-----CC 22632
QY 861 oValLysSerGlu-----CysThrGluGluAlaGluGluGlyProAlaLysGlyLy 878
```

```

FT /label= epoE_gene
FT /note= "encodes modules 7 and 8"
FT 38014. .39375
FT /*tag= az
FT /note= "encodes KS7"
FT 39589. .40626
FT /*tag= ba
FT /note= "encodes AT7"
FT 41341. .41922
FT /*tag= bb
FT /note= "encodes KR7"

Alignment Scores:
Pred. NO.: 5.73e-10 Length: 71989
Score: 622.50 Matches: 660
Percent Similarity: 31.32% Conservative: 265
Best Local Similarity: 22.35% Mismatches: 1065
Query Match: 4.71% Indels: 970
DB: 3 Gaps: 145

US-09-522-753-5 (1-2517) x AAA29349 (1-71989)
QY 32 ArgThrHisThrAspValGlyLeuLeuGlu-----TyrGlnHisHisSerArgAsp 48
DB 25082 AAGCGGACGTCACGCTGAGGAGTTGGGATCTCCACGCCATCAACGAGTC 25023
QY 49 TyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgProSerLeu 68
DB 25022 GAGGCGCTGCTCATTAACCCCATCCAGGATCGAGCGCTCGGGT-CGAGGAAGCCCA 24964
QY 69 LeuSerGluPheGlnPro---GlyAsnGluArgSerGlnGluLeuHisLeuArgProGlu 87
DB 24963 GCACCGGAGCAGCGCCCGCCATGAGAGCCTCTGTCAGAGCCATCGGGCTT---CCGCAA 24907
QY 88 -----SerHisSerTyrLeuProGluLeuGlyHisSerGluMetGluPhe 102
DB 24906 CGGACAGCCACGCCAGTTCGGGTGTCGCGCGGAGGGGAAGGCG-----CGATGA 24853
QY 103 IleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerPro 122
DB 24852 TGTGCGCGCTCGGACCGCCGAAACGAGTTCGCGAGCGCCCTCGAGCGGTATACCG 24793
QY 123 LeuLeuAlaThrGlyGlnProLysLeuAspLeuThrLysAspArgSerLeuThr 142
DB 24792 GCGGACAGCGGCCGATCCATCCGGGTCACTGCGGTGAGCGCGCGCTCTCCACCA 24733
QY 143 GlyLysLeuGluProValSer----- 149
DB 24732 GCGCGTGGAGCGCGACAAACGTCGCGAGCTGACATGGGCAGAACCCCGATGTCGTCA 24673
QY 150 -----ProProSerProPro-----HisThrAspProGluLeuGlu 161
DB 24672 GACGTGCATGAGCCTCCGCGTCCGCATGCTCCCTCGGCCACAGACCCACCGGAGCG 24613
QY 162 LeuValPro-ProArgLeu-----SerLysGluGluLeuIleGlnAsnMetAs 177
DB 24612 TCACGCGAGCGAGATTCGACACCGCCAGNAGTGGCGAGCCGTCGAGGAAGCGTTGG 24553
QY 177 pArgValAspArgGluIleThrMetValGlnGlnIleSerLysLeuLysLysGlu 197
DB 24552 CGCGCGGTACGACCTTGCTATGGTATACCCACACCGCTCGCGCCGACGAGAACACA 24493
QY 197 nGlnGlnLeuGluGluAlaAlaLysProGluProGluLysProValSerProPr 217
DB 24492 CGAACAGGTTCGAGAGCGCGCGTGCAGCAGCGGTGACGAG-----CCAGC 24445
QY 217 oProlleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLy 237
DB 24444 TCCT-----GGCCA 24436
QY 237 sAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTy 257
DB 24435 CCTTGGGACGAGACCGACTCGAGCAGGGTCTC-----GTCCGCTCTCCGCGAGTG 24385

```

```

QY 257 rAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetAr 277
DB 24384 GAGCGATGAGCTGAGCCAGCGGGCTGCAC-----CACCCTCGCAGCG 24340
QY 277 gLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLy 297
DB 24339 GGG-----CTGACCGACGAAACACGCGC----- 24315
QY 297 sPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysValGluArgIleGl 317
DB 24314 ----TGTATCGGTTCGAC-----GTCGGCCAGCTCCACGCTGCCA 24277
QY 317 uAsnAsnProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPh 337
DB 24276 CGGTACCCCGTGCACC----- 24261
QY 337 eProGluIleArgLysGlnArgGluLeuGlnArgMetGlnSerArg----- 353
DB 24260 -CGCGCCTCCAGCGCTCGACCGCTGCGATCCGCGCGGGATCTCAGGCGGTGTGTCT 24202
QY 354 -----ValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSe 368
DB 24201 CGGCCACGCTGCGGTGCGGCACCCCGCGCGGTGTCAGCACCAAGTCCGCGCTC 24142
QY 368 rGlu-----HisGluValSerGluIleIleAspGlyLeuSerGluGl 382
DB 24141 CAGCTCCACCGACCTGCGGCACCATCAGGCCCGACGCCCGAGGCTCCCGTCACCA 24082
QY 382 nGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAl 402
DB 24081 GGTAGTCTGCGCTCCGACGACGCGCGGTGCGCTCTCCCGCGGGGGGGGCCCA 24022
QY 402 aAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422
DB 24021 CCAGCGTCCGCGGTG-----CCGCGGCCCAT 23995
QY 422 lTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe----- 440
DB 23994 GCGGAAGGCGAGTGTCTCTGTCGACAAATAGCTCGGTGACGAGCATCTCGC 23935
QY 441 -----ArgGluLysPheMetGlnHisProLysAsnPh 451
DB 23934 CGTGCATGCGGTGCTGGGCGGCTCGCTCGGGATCCAGGTCCAGCGCCGCC 23875
QY 451 eGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTy 471
DB 23874 AGGC-----CCCGGGATGCTCGAGCGCCGCCCGGCCCA-TGCCCCATA----- 23830
QY 471 rLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGlu 491
DB 23829 -----ACGCGCGCTGACAAGGG 23813
QY 491 yLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPr 511
DB 23812 GCGATCG-----CAGCTCGTCCCAACCATG 23786
QY 511 oMetProArgSerSerGln-GluGlu-----LysAspGluLysGluLysGluLysGluA 529
DB 23785 CATCCCCCGGTTCACGACCCAGAGTCGGGGCGAACAGACACGGTGTCTCAGAAACCGA 23726
QY 529 laGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLysGluL 549
DB 23725 GCCAAGCGAGCACCGCGCGGTAGCACGCGGTGCGCGCATCTCATCATCGAC 23666
QY 549 ysThrAspAspThrSerGlyGluAspAsnAspGluLysGlu-----AlaValAlas 566
DB 23665 GCCTCCGCGCGCAGCAGCGCGCTCCAGACCCACAGGTAGACACTACTCTGCAATCGCTT 23606
QY 566 erLysGlyArgLysThrAlaAsnSerGlnGlyArgArg-----LysGlyArgIleT 583
DB 23605 CG-----ACCGCGGCGCTCGGTCCACGAGCTCGGCGGTGCGCGATGTCTCT 23558

```

```
FT FT /*tag= e
FT FT /note= "encodes acyl transferase (AT) of the loading
FT FT domain"
FT FT 4917. .5810
FT FT /*tag= f
FT FT /note= "encodes enoyl reductase (ER) of the loading
FT FT domain, potentially involved in formation of the thiazole
FT FT moiety"
FT FT 5856. .6155
FT FT /*tag= g
FT FT /note= "encodes acyl carrier protein (ACP) of the loading
FT FT domain"
FT FT 6260. .10493
FT FT /*tag= h
FT FT /label= epoB gene
FT FT /note= "encodes module 1, the NRPS module"
FT FT 6861. .6887
FT FT /*tag= k
FT FT /note= "encodes heterocyclization signature sequence"
FT FT 6861. .6887
FT FT /*tag= l
FT FT /note= "encodes condensation domain C4 of the NRPS
FT FT module"
FT FT 7358. .7366
FT FT /*tag= m
FT FT /note= "encodes condensation domain C7 (partial) of the
FT FT NRPS module"
FT FT 7898. .7921
FT FT /*tag= n
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT 7898. .7921
FT FT /*tag= o
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT 8261. .8308
FT FT /*tag= p
FT FT /note= "encodes adenylation domain A3 of the NRPS module"
FT FT 8411. .8422
FT FT /*tag= q
FT FT /note= "encodes adenylation domain A4 of the NRPS module"
FT FT 8861. .8905
FT FT /*tag= r
FT FT /note= "encodes adenylation domain A6 of the NRPS module"
FT FT 8966. .8983
FT FT /*tag= s
FT FT /note= "encodes adenylation domain A7 of the NRPS module"
FT FT 9090. .9179
FT FT /*tag= t
FT FT /note= "encodes adenylation domain A8 of the NRPS module"
FT FT 9183. .9992
FT FT /*tag= u
FT FT /note= "encodes oxidation region for forming thiazole"
FT FT 10121. .10138
FT FT /*tag= v
FT FT /note= "encodes adenylation domain A10 of the NRPS
FT FT module"
FT FT 10261. .10306
FT FT /*tag= w
FT FT /note= "encodes thiolation domain (PCP) of the NRPS
FT FT module"
FT FT 10839. .16137
FT FT /*tag= x
FT FT /label= epoC gene
FT FT /note= "encodes module 2"
FT FT 10654. .12033
FT FT /*tag= y
FT FT /note= "encodes KS2, the KS domain of module 2"
FT FT 12250. .13287
FT FT /*tag= z
FT FT /note= "encodes AT2, the AT domain of module 2"
FT FT 13327. .13899
FT FT /*tag= aa
FT FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT FT module 2"
FT FT
```

```
FT FT misc_RNA
FT FT 14962. .15756
FT FT /*tag= ab
FT FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT FT module 2"
FT FT 15763. .16008
FT FT /*tag= ac
FT FT /note= "encodes ACP2, the ACP domain of module 2"
FT FT 16134. .37907
FT FT /*tag= ad
FT FT /label= epoD gene
FT FT /note= "encodes modules 3-6"
FT FT 16425. .17606
FT FT /*tag= ae
FT FT /note= "encodes KS3"
FT FT 17817. .18857
FT FT /*tag= af
FT FT /note= "encodes AT3"
FT FT 19581. .20396
FT FT /*tag= ag
FT FT /note= "encodes KR3"
FT FT 20424. .20642
FT FT /*tag= ah
FT FT /note= "encodes ACP3"
FT FT 20706. .22082
FT FT /*tag= ai
FT FT /note= "encodes KS4"
FT FT 22296. .23336
FT FT /*tag= aj
FT FT /note= "encodes AT4"
FT FT 24069. .24647
FT FT /*tag= ak
FT FT /note= "encodes KR4"
FT FT 24867. .25151
FT FT /*tag= al
FT FT /note= "encodes ACP4"
FT FT 25203. .26576
FT FT /*tag= am
FT FT /note= "encodes KS5"
FT FT 26793. .27883
FT FT /*tag= an
FT FT /note= "encodes AT5"
FT FT 27966. .28574
FT FT /*tag= ao
FT FT /note= "encodes DH5"
FT FT 29433. .30287
FT FT /*tag= ap
FT FT /note= "encodes ER5"
FT FT 30321. .30869
FT FT /*tag= aq
FT FT /note= "encodes KR5"
FT FT 31077. .31373
FT FT /*tag= ar
FT FT /note= "encodes ACP5"
FT FT 31440. .32807
FT FT /*tag= as
FT FT /note= "encodes KS6"
FT FT 33018. .34067
FT FT /*tag= at
FT FT /note= "encodes AT6"
FT FT 34107. .34676
FT FT /*tag= au
FT FT /note= "encodes DH6"
FT FT 35760. .36641
FT FT /*tag= av
FT FT /note= "encodes ER6"
FT FT 36705. .37256
FT FT /*tag= aw
FT FT /note= "encodes KR6"
FT FT 37470. .37769
FT FT /*tag= ax
FT FT /note= "encodes ACP6"
FT FT 37912. .49308
FT FT /*tag= ay
FT FT
```

| | | | | |
|------------|-------------|--|--|------|
| QY | 2407 | GlyArg | -----ProSerSerArgLysAlaLysSer | 2416 |
| Db | 19843 | AGGAACCAATCGTCTTCTTCGCGCGCGGCACCCCTCCGGGAAGCGCTGCCACGAGC | 19784 | |
| QY | 2417 | Pro | -----AlaProGlyLeuAlaSerGlyAspArg | 2429 |
| Db | 19783 | CGCAAACTTCGGCGACCACTGCACCCGAGCTGTGCACCAACCAAGTCGGCGCCTGC | 19724 | |
| QY | 2430 | ValSerSerVal | -----HisSer | 2436 |
| Db | 19723 | CGATCGGGGGTTTGGTGCCCATGGTTTACAGCGCGGCATGGCACCACAGTCCCCCGAA | 19664 | |
| QY | 2437 | GlyAspCysAsnArg | -----ThrPro | 2444 |
| Db | 19663 | GGCGCTGCAAGAGCCGCGAGCGCCCACTCCACGGGCACCCACGCGCTCGCTCGCCA | 19604 | |
| QY | 2445 | LeuThrAsnArg | -----ValTrpGluAspArgProSerSerAlaGlySerThr | 2460 |
| Db | 19603 | CTGCGGGCGAAGAGGCTGCCAGCATCTGGAAGCAGCGTCCAGCAGCGCAGGATGCAC | 19544 | |
| QY | 2461 | ProPhe | -----ProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro | 2479 |
| Db | 19543 | CGATACTCCGTCGCGAGCGCGCGCTCCGCGCGGCGTACCCGTCGCCAGCGCCTCGCT | 19484 | |
| QY | 2480 | ProPro | -----ProGlyLeuProAlaGlySerGlyProLeuAlaGlyPro | 2494 |
| Db | 19483 | TCACCCCGCATAGTCTCAGCAATCCCTCGAAGCGAGCGCGCTACTGCGAGCCCC | 19430 | |
| RESULT 66 | | | | |
| AAA29349/c | | | | |
| ID | AAA29349 | standard; DNA; 71989 BP. | | |
| XX | AAA29349; | | | |
| XX | 15-SEP-2003 | (revised) | | |
| DT | 12-SEP-2000 | (first entry) | | |
| XX | | | | |
| DE | | Sorangium cellulosum epothilone polyketide synthase operon genomic DNA. | | |
| XX | | | | |
| KW | | Epothilone; polyketide synthase; epaA; epoB; epoC; epoD; epoE; epoF; | | |
| KW | | epoL; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal; | | |
| KW | | tubulin polymerization assay; anti-tumour; cytostatic; ds. | | |
| XX | | | | |
| OS | | Polyangium cellulosum. | | |
| XX | | | | |
| PH | | Location/Qualifiers | | |
| FT | | 3..992 | | |
| FT | | /tag= a | | |
| FT | | /label= ORF A | | |
| FT | | /product= "transposase" | | |
| FT | | /note= "not part of the PKS" | | |
| FT | | 989..1501 | | |
| FT | | /tag= b | | |
| FT | | /label= ORF B | | |
| FT | | /product= "transposase" | | |
| FT | | /note= "not part of the PKS" | | |
| FT | | 1998..6263 | | |
| FT | | /tag= c | | |
| FT | | /label= epaA gene | | |
| FT | | /note= "encodes the loading domain" | | |
| FT | misc_RNA | 2031..3548 | | |
| FT | | /tag= d | | |
| FT | | /note= "encodes ketide synthase (KS-Y) of the loading domain" | | |
| FT | misc_RNA | 2031..3548 | | |
| FT | | /tag= i | | |
| FT | | /note= "encodes condensation domain C2 of the NRPS module" | | |
| FT | misc_RNA | 2031..3548 | | |
| FT | | /tag= j | | |
| FT | | /note= "encodes condensation domain C2 of the NRPS module" | | |
| FT | misc_RNA | 3621..4661 | | |

| | | | | | | | |
|----|-------|---|-------|----|-------|---|-------|
| QY | 1041 | sLeuProGlyAspPro-----ProCys-- | 1048 | QY | 1296 | erSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluG | 1316 |
| Db | 24927 | CCACCGTGCAGGCCCTCGGTCACGATCAGAGCGGGGTGAGCGCGCGCTGCGG | 24868 | Db | 23852 | CGGCTCGGGCGCTCCGAGCGACACCATCGCGGCC--CGCCGAGAGCCCTGTCATCAGC | 23794 |
| QY | 1049 | -----TyrThrSerGI | 1052 | QY | 1316 | lyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleP | 1336 |
| Db | 24867 | AGCGCTGAATCAGCGCAGACACCGCGCGCGCAAGATGGGTGACTTTGGGACCTCT | 24808 | Db | 23793 | CG-----C | 23791 |
| QY | 1052 | yLeuProPheProValProArgGluValIleLysAlaSerProHisAlaProAspPr | 1072 | QY | 1336 | roProGluArgHisSerProHisHisLysLysGluLmHisHisIleArgGlySerIleT | 1356 |
| Db | 24807 | TCGGCCCATGCCCGCGCTCCAGACGGCGTCCAGACCCACAGACAGTACAGACCCCTGC | 24748 | Db | 23790 | CGCGCGCGCGCACAGAGCTTACCCATCTCCACGAGAGACACCCCGCCAGCAGC-C | 23732 |
| QY | 1072 | oSerAlaPheSerTyrAlaProProGlyHisProLeu-----ProLeuGI | 1087 | QY | 1356 | hrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaL | 1376 |
| Db | 24747 | CAGTCGTGGCGCCACCGAGGCGCTGGGTCACTCTCGCAACCGCGAGGCTTCGGCG | 24688 | Db | 23731 | CGCCACCACTCCCGATGCTATGCCAACACAGGAGCTCCGGCTCTAGCCGCCACAGCG | 23672 |
| QY | 1087 | yLeuHisAspThrAlaArg-----ProValLeuProArgProProThrIleSerAs | 1104 | QY | 1376 | ysLeuLeuLysArgGluGlyThrProPro----- | 1385 |
| Db | 24687 | GGCGCATGGACACGGCGGCAGAACATCCCTTGCAGCAAGCGCGCGG-CGCCCTC | 24629 | Db | 23671 | CCACAGCGCGTACGCGGTACTCCACCGCGAAGAGCGGGCTGGTGAACCGCGTCTG | 23612 |
| QY | 1104 | nPro-----ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGI | 1123 | QY | 1386 | -----ProProProProSerArgAspLeuThrGluAlaT | 1397 |
| Db | 24628 | CCCACTCCACCCCGTGGCGGCAGACAGCACCCCGCGAGCGGCTCGCGCGAATC | 24569 | Db | 23611 | GTGAGCAACAGACTCGCGCTCCCGCTCCCGCCACATCA----- | 23568 |
| QY | 1123 | yAlaIleSerGlnGlyMetSerValGln-----LeuHisValProTyrSerGluHisAlaLy | 1142 | QY | 1397 | ryIleThrGlnAlaLeuGlyProLeuLysLysLysProAlaHisGlu----- | 1412 |
| Db | 24568 | CACGGATGAGCGGATCTCGGGCCATCCACCCGGTAGAACCACTGCGCGCAGCGATC | 24509 | Db | 23567 | --CCTCGCGAGCGGGCTCCAGTCTCCCGGTCCGACAGCGCACACCGGTGCAACGC | 23510 |
| QY | 1142 | sAlaProValGlyProValThrMetGly-----LeuProLeuPr | 1155 | QY | 1413 | -----GlyLeuValAlaThrValLysGluAlaGlyArgSerIle----- | 1425 |
| Db | 24508 | GGCGCGCTGGCTCGAG--CCCTCGCGCGCGGCTCGATCCAGTACCGTTCGCGTCC | 24450 | Db | 23509 | CTCCCGAACGCTGGCGCACCGCGCGCAAGAGCCCG--CGGCCATGCCCGCGTCTGCGCG | 23451 |
| QY | 1155 | o---MetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLmLeuSerPr | 1174 | QY | 1426 | -----HisGluIleProArgGluGluLeuArgHisThrProGluL | 1439 |
| Db | 24449 | ACGGATAGTCCGACAGCGCACCGCGCCGCGAGCGGTGGGAAGACGCCCGCGAGCTGA | 24390 | Db | 23450 | CTGTCCGTGAACAGAACCGCAGCTTCGCGCGAGAGCTCGCGATCGAGCGCGCG | 23391 |
| QY | 1174 | oArgGlyGlnAlaGlyPro-----ProGluSerLeuGlyVa | 1186 | QY | 1439 | euProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysT | 1459 |
| Db | 24389 | CGAGCGCGCGCGGCCACACCTCCCGCGCGCTCGAGCGCGCTCGAGCACCCCGCGACCTCTCTCGC | 24330 | Db | 23390 | CCCGCGCGCGTCTGCTCGCGCACCGCGCGAGAGCGCGCACAGCGCCCTCGCGCG | 23331 |
| QY | 1186 | ProThrAlaGln--GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGly- | 1205 | QY | 1459 | yrAspThrGlyAlaSerThrThr--GlySerLysLysHisAspValArgSerLeuIleGly | 1478 |
| Db | 24329 | CGCCCGCGCGAACAGCGCACCGACCGTC--GGTCCGCTCCCGCAGGCGAGGTGGCA | 24273 | Db | 23330 | ACGTACCGCACCGGAGCGGTGCTCATCGCGC-----TCGCGGTGCG | 23286 |
| QY | 1206 | -----GlySerIleThrLysGlyIleProSerThrArgValPro-----SerA | 1220 | QY | 1479 | SerProGlyArgThr----- | 1483 |
| Db | 24272 | ACAGCCGAGCAGCGTGGCTTCGGGCGCCACTTCGACGAACGTCCCGCGCGGCTTCGT | 24213 | Db | 23285 | TCGCCAGCTGAACCGCACCGTCCCGAGCCCGAGCTCCGGTGCATGTCCAGGTGTGCG | 23226 |
| QY | 1220 | spSerAlaIleThr-----TyrA | 1226 | QY | 1484 | -----PheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArg | 1501 |
| Db | 24212 | GCAGCGCTTACCCCGTCCGGAAGCGCACCGCTCCCGCACGTGCGCACCCAGTACC | 24153 | Db | 23225 | GCAGCGCGCGCTTCGCGCAT-----CGAGCGCCCTCGC | 23190 |
| QY | 1226 | rgGlySerIleThrHisGlyThr-----ProAlaAspValLeuTyrLysGlyThrI | 1243 | QY | 1502 | AlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySer | 1521 |
| Db | 24152 | CGGGCGCTCAGTCTCGTGCAGCACCTTCCTCGCTCAGTGTGCTCACCAGCGAAACGC | 24093 | Db | 23189 | TCITTC-----CCGACAGCACAAAGCTCCCGAGCGCTCCGGCGCGGAGCCACA | 23136 |
| QY | 1243 | leThrArg-----I | 1246 | QY | 1522 | IleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeu | 1541 |
| Db | 24092 | TTGGCGCGGTACGTACCGACGCGCGCCACCGCCGCAACTCTCTCCAGATCGGTCCA | 24033 | Db | 23135 | GCTCACCGCGCGCTCTTCAGCACCATCATCGCGTTCGTTCCTCCCTATCCCGAACG | 23076 |
| QY | 1246 | leIleGlyGlu-----AspSerProSerArgLeuAspArgGlyArgGluAsps | 1262 | QY | 1542 | ThrTyr-----GluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro | 1559 |
| Db | 24032 | TCAGCGCGAGTGGAAACGGTGGAGACATGACGCGCTTGGTGGCGCGCGCGCGCG | 23973 | Db | 23075 | AGCTACACCCCGCAAGCGCGCGGTCCGTGCGCGCCACCGCACCGGTCTCGTCCCA | 23016 |
| QY | 1262 | erLeuProLysGlyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluG | 1282 | QY | 1560 | ValThrMetArg-----GluProThrProArgLeuGlnGluGlySerLeuSerSerSe | 1577 |
| Db | 23972 | CGAACCCCGCGGATCGCTGACCGCTTGTCTCCACGCGCGCGATCAACCTGCTCG | 23913 | Db | 23015 | ACCGAGCGCGCTGCTCCCTCGAGCGCGATTCGCGATTGAGCGTGGGAAGTTGAGTTTC | 22956 |
| QY | 1282 | lyGlyMetSerVal-----ThrGlnCys---SerLysGluAspGlyArgS | 1296 | QY | 1577 | rLys-----AlaSerGlnAspArgLysLeuThrSerThrProArg | 1590 |
| Db | 23912 | GCACATTGACCGCGGATCGACACCGAGCGCGCTGGCGCGCCACCGCGCGCCACT | 23853 | Db | 22955 | TCGGATCGCTCGTGTGCTCAGCAAGCGGTGCTTGTATCAGGCGCGCTAGCGTCCG | 22896 |
| | | | | QY | 1591 | -----GluIleAla-----LysSerProHisSer-----ThrValPr | 1601 |


```
Db 26886 CGTCGAAGCTCTGGCAGCGAG----- 26864
QY 478 snTyrLysSerLeuValArgSerTyrArgArgArgGlyLysSerGlnGlnGlnG 498
Db 26863 -----GTCCAGCGC---CACGAGCAGAGA---GCACGCCGTATCCACA 26824
QY 498 lngInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 518
Db 26823 GCCAGGTCGGCGCTGCAGCGCCAGGAAAATGATAGCGCTCCCGCGCAACGCTGAGC 26764
QY 518 luGluLysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLys 538
Db 26763 AT----- 26762
QY 538 alGluAsnAspLysGluAspLeuLysGluLysThrAspThrSerGlyGluAspA 558
Db 26761 -----GTTGCCGCTGCCGCTG 26746
QY 558 snAspGluLysGluAlaVal-AlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArg 577
Db 26745 TAGAGCCCCCGCTCATCGCGGAGTTCCTGCACCGCTCGCATATTCTGTTGGGGCCC 26686
QY 578 ArgLysGlyArgLysThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThr 597
Db 26685 -GCGGCC-----ACGAACAGCCCGTGGGGCTCTCGCGCAGCGCGCGGTCCTCG 26635
QY 598 ProGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTTPThrGlu 617
Db 26634 CCCCGCGCTCG-----ATCGCTCCACGTCATCTTCACGACGACCGCTGTTGGGG 26581
QY 618 GluGluMetGluThrAlaLysLysGlyLeuGluGlnHisGlyArgAsnTrpSerAlaIle 637
Db 26580 TCCAGGCTC-----ATCGCTCCCGAGCGAGATGTGGAAGAGGCC 26539
QY 638 AlaArgMetValGlySerLys----- 644
Db 26538 GCATCGAAGCTCTCCACCTCGCGCAGAAAGCCACCCCTGGGCACGTAGTCTGTCTGT 26479
QY 644 ----- 644
Db 26478 GCCTCTCCGAGCGGGGAGCGCGCTGTGCCCATTCACCGGTGCGCGGACCTCG 26419
QY 645 -----ThrValSerGlnCysLys-----AnPheTyrPheAsn 655
Db 26418 GTGCTGACCACCGCCCTCGGTCAACAGCTGCAGTAGGACTCCAGGTCTCGACCCCG 26359
QY 656 TyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLysMetGluLys 675
Db 26358 CCGGGAAGCGGAGCGGCGCACCCACGATGG--CGATCGGGTCTCTGTGCGCGCAACC 26301
QY 676 GluArgAsnAlaArgArgLysLysLysLysLysLysLysLysLysLysLysLys 695
Db 26300 GAAGCTCCGACGTCGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 26247
QY 696 PheProProValVal-----GluAspGluGluMetGluAlaSerGlyValSerGlyAsn 713
Db 26246 ATCCACCGCGCTCCACGCTCGATGTGCGAAGCTAGGCTGCGCGCACGCGGCATAC 26187
QY 714 GluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValPro--- 732
Db 26186 CCAGCTCACCGCTGAAGCGGTTTGGGATCTCCAGCGCATCAGGAGTCGAGGCGCTGCT 26127
QY 733 -----ArgGlyGluCysSerGly----- 738
Db 26126 CGGCGAAGCTCTGTGTCAGCGTGCAGCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTG 26067
QY 739 -----ProAlaThrValAsnAsnSerSerAspThrGluSerIle 751
Db 26066 CCAGCGCGCACAGCTCGTAGAGCGCGGAGGCTCTCCACACAGACGCGTGTGGCC 26007
QY 752 ProSerProHisThrGluAlaAlaLysAsp-----ThrGlyGlnAsnGlyProLys 768
```

| | | | |
|----|-------|--|-------|
| Qy | 13 | ArgAlaThrGluProArgTyrProProHisSer-LeuSerTyrProValGlnIleAlaAr | 32 |
| Db | 28316 | CGAGCGACACCATCGCGCGCGCGGAGAGCCCTGTCATCAGCGCGCC | 28263 |
| Qy | 32 | gThrHisThrAspValGlyLeuLeuGluTyrGlnHis-----HisSerArgAspTyrAl | 50 |
| Db | 28262 | CCGCCACGAGCCTCACCCCATCTTCAGCGAGAACACCCCGCCACGCACGC | 28211 |
| Qy | 50 | aSerHisLeuSerProGlySerIlelleGlnProGlnArgArgProSerLeuLeuSe | 70 |
| Db | 28210 | -CGCCACCAAGCTCCCGGGGTATGTCGCCAGCCACCAAGCTCGCG-----CTCTA | 28164 |
| Qy | 70 | rGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSe | 90 |
| Db | 28163 | CGCCCCACGACCGCCACAGCGCGCTCAGCGGTACTCCACCGTGA----- | 28118 |
| Qy | 90 | rTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLe | 110 |
| Db | 28117 | -----GAGCGCGGCTGGGTGAACCGCGTCTGTGTCGAGCAACACGA----- | 28076 |
| Qy | 110 | uGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAl | 130 |
| Db | 28075 | -----CTCGGCGCTCCCGGCTCGCCACATCACTTCGC-GCAGCGGGCGGTCCAGC | 28024 |
| Qy | 130 | aGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerPr | 150 |
| Db | 28023 | T-----CCCGGTCCGAACAGCGCCACGCACCGCGG | 27997 |
| Qy | 150 | oProSerProProHisThrAspProGluLeuGluLeuValProProArgLeuSerLysgl | 170 |
| Db | 27996 | TCGAACGCGCTCCCGGAACG-----CTGGCCACGCCCGCGCAAGACCCCGCG | 27952 |
| Qy | 170 | uGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIl | 190 |
| Db | 27951 | CCCATGCCCGCGTCTGCG-----CGCCCTGTCCGGTGAACAGSACGCCACGC | 27904 |

```
Qy 2382 ProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGly 2401
Db 38166 CCCACGGTTCGCCCGACGCTGATTATGTCGTCGCCGCACTACCGCGGTGCCGCCGC 38225
Qy 2402 -----LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSer 2416
Db 38226 GCCGACACCCCGCCAGACCGCCATCCCGCGCGCGCGCGCCGCAAGATCCGGAATCC 38285
Qy 2417 ProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSer--ValHisSerG 2436
Db 38286 GCCGCGCCCGCCCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38345
Qy 2436 luGlyAspCysAsnArgThrProLeuThrAsnArgValTrpGluAspArgProSerS 2456
Db 38346 CCGCGCGGTGCGCGCGTGGTTGCC-----CGCGG-CGCGCCACGACCGCGCGGT 38395
Qy 2456 erAlaGlySerThrProPheProTyAsnProLeuIleMetArgLeuGlnAlaGlyValM 2476
Db 38396 CGCGGACAGCGCGCGCTTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38455
Qy 2476 etAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisH 2496
Db 38456 CTGCGAATCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38515
Qy 2496 ieAlaTrpAspGluGluProLysProLeuLeu 2506
Db 38516 CGCGCGCGCGCGCGCGCGTCCCGTCCCGTGTG 38547

RESULT 65
AAZ55887/C
ID AAZ55887 standard; DNA; 68750 BP.
XX
AC AAZ55887;
XX
DT 15-SEP-2003 (revised)
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Polyangium cellulosum.
XX
FH Key
FH 1. .1826
FH CDS
FH /tag= a
FH /partial
FH /product= "Partial Orf 1 protein (AAZ55880)"
FH /note= "No initiation codon given in the specification"
FH complement(1900..3171)
FH /tag= b
FH /product= "Orf 2 protein (AAZ55881)"
FH 3415. .5556
FH CDS
FH /tag= c
FH /product= "Orf 3 protein (AAZ55882)"
FH complement(5612..5992)
FH /tag= d
FH /product= "Orf 4 protein (AAZ55883)"
FH 5226. .6675
FH CDS
FH /tag= e
FH /product= "Orf 5 protein (AAZ55884)"
FH 7610. .11875
FH CDS
FH /tag= f
FH /product= "Type I polyketide synthase, EPOS A (AAZ55873)"
FH 11872. .116104
FH CDS
FH /tag= g
FH /product= "Non-ribosomal peptide synthetase, EPOS P
FH (AAZ55874)"
FH 16251. .21749
FH CDS
FH /tag= h
FH /product= "Type I polyketide synthase, EPOS B (AAZ55875)"
FH 21746. .43519
```

```
FT /tag= i
FT /product= "Type I polyketide synthase, EPOS C (AAZ55876)"
FT 43524. .54920
FT CDS
FT /tag= j
FT /product= "Type I polyketide synthase, EPOS D (AAZ55877)"
FT 54935. .62254
FT CDS
FT /tag= k
FT /product= "Type I polyketide synthase, EPOS E (AAZ55878)"
FT 62369. .63628
FT CDS
FT /tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT (AAZ55879)"
FT 63779. .64333
FT CDS
FT /tag= m
FT /product= "Orf 6 protein (AAZ55885)"
FT complement(63853..64290)
FT CDS
FT /tag= n
FT /product= "Orf 7 protein (AAZ55886)"
FT complement(64287..64727)
FT CDS
FT /tag= p
FT /product= "Orf 9 protein (AAZ55888)"
FT 64363. .64920
FT CDS
FT /tag= o
FT /product= "Orf 8 protein (AAZ55887)"
FT complement(65008..65874)
FT CDS
FT /tag= r
FT /product= "Orf 11 protein (AAZ55890)"
FT 65063. .65767
FT CDS
FT /tag= q
FT /product= "Orf 10 protein"
FT complement(65871..66338)
FT CDS
FT /tag= s
FT /product= "Orf 12 protein (AAZ55891)"
FT 66667. .67137
FT CDS
FT /tag= t
FT /product= "Orf 13 protein (AAZ55892)"
FT 67334. .68251
FT CDS
FT /tag= u
FT /product= "Orf 14 protein (AAZ55893)"
FT 68346. .68750
FT CDS
FT /tag= v
FT /partial
FT /product= "Partial Orf 15 protein (AAZ55894)"
FT /note= "No termination codon given in the specification"
```

W09966028-A2.

23-DEC-1999.

16-JUN-1999; 99WO-EP004171.

18-JUN-1998; 98US-00099504.

24-SEP-1998; 98US-0101631P.

05-FEB-1999; 99US-0118906P.

(NOVS) NOVARTIS AG.

(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;

WPI; 2000-097741/08.

P-P5DB; AAZ55873, AAZ55874, AAZ55875, AAZ55876, AAZ55877, AAZ55878,

AAZ55879, AAZ55880, AAZ55881, AAZ55882, AAZ55883, AAZ55884, AAZ55885,

AAZ55886, AAZ55887, AAZ55888, AAZ55889, AAZ55890, AAZ55891, AAZ55892,

AAZ55893.

New isolated epothilone synthase genes, used for the recombinant

production of epothilone for use in cancer therapy.

Claim 14; Page 87-104; 174pp; English.

This sequence represents a 68.75 kb contig from Sorangium cellulosum

comprising 22 open reading frames (ORFs) and includes genes encoding

Db 36449 -----TACCGGTTGAACCCGGTGTCTTTGAAGGACACCGCCGAGGCTGATCTGGGG 36501
Qy 1712 YLeuSerProArgGluSerSerLeuAlaLeuAenTyraAlaAlaGly--ProArgGlylle 1731
Db 36502 GATCTGGCGCGCGCCGACCATGGCTGGCGGCTGTGTCTGTATGGCGGTTCGCGGTAC 36561
Qy 1732 IleAspLeuSerGlnVal---ProHisLeuProValLeuValProProThrProGlyThr 1750
Db 36562 AGCCCGTTGAGCCCGTCTCGCGATGATCAGCCCTGTGTGGCGCGCGCGGTGTG 36621
Qy 1751 ProAlaThrAlaMetAspArgLeuAlaTyLeuProThrAlaProGlnProPheSerSer 1770
Db 36622 CCGCCG-----GCTTCCCGCGGTTCGCGCGGTTCGCG----- 36654
Qy 1771 ArgHisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThr 1790
Db 36655 -----CCGTTCGCGATCAG-----ACGGCGTTGGGGACC 36684
Qy 1791 -ThrSerSerSerGluArgGluArgAspArgGluArgGluArgAspArgGlu-A 1810
Db 36685 GAGCTCGAATTCACACCGGTTCAGCGCCAAACCCGCGCGCGCGCGCTCGCGCGCTTG 36744
Qy 1810 xGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProG 1830
Db 36745 CCGNACACCCCGCGGTACCGCGTACCGCC-----G 36777
Qy 1830 lYThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSera 1850
Db 36778 GTCCCGCGCTGTCTGCGATGCGCTGAGCCACCGGTGCGCGCGCACCGCGAAGCCG 36837
Qy 1850 rgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg-ThrGlnAsp 1869
Db 36838 AAGACGAGCGCCGACCTCCGCGCTTCCCGCGACCCCGCGGTTCGCGCGACATTAAAG 36897
Qy 1870 AlaLeuGlnGlnArgProSerValLeuHisAsnThr-----GlyMetLysGlylle 1887
Db 36898 GCGTCCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGGCGAGTAGCGCGGTG 36957
Qy 1888 ThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProVal 1907
Db 36958 CCGCGCGTCCGCGCGGTTCGCGCGAAGGTGCGCGCTGTGCGCGCACCGCGCAGTG 37017
Qy 1908 ArgProAlaAlaThrPheProAlaThrHisCysProLeuGlyGlyThrLeuAspGly 1927
Db 37018 CCGCGCGCGCGAACGCCCGCG-----TGCCCC----- 37047
Qy 1928 ValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArg 1947
Db 37048 -----CCGCGC-----CCGCGCTGCGCGCGCGCGCGCGTCCCGCC 37080
Qy 1948 ProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProPro---AlaArg 1966
Db 37081 CCGCGGTGCGCGCGCGCGCGAAGAGCAATCCGTTCGCCCGGTCCCGCCATTCCGCCCA 37140
Qy 1967 SerGlyLeuGluProAlaSerSerProSerLysGlySer-----GluProArgPro 1983
Db 37141 AACCCCG 37200
Qy 1984 LeuValProProValSerGlyHisAlaThrIle-----AlaArgThrPro 1998
Db 37201 CCGGTCCCGCGC-----GCACCGATAAGTTTTCGGAGAGCGCGCGCTGGCCG 37248
Qy 1999 AlaLysAsnLeuAlaProHisHis-----AlaSerProAspProAlaProPro 2015
Db 37249 CCGGTCCCTGGGACCCAGAGAACAGCGCGGTTCACCGCGCGCGCGCGCGCGCGCGCG 37308
Qy 2016 AlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGlu 2035
Db 37309 GTGTCCAGGCCAAACCCCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37356
Qy 2036 LeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerProGluGlyValGluPro 2055
Db 37357 -----AGCAAGCGCGGTTCGCCCGTTCGCCCG 37383

Qy 2056 ValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlu 2075
Db 37384 GTCCCGCTTGGCCCGTTCGATGCGCGGACCCCGCGGTGCGCCCA----- 37434
Qy 2076 GluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys 2095
Db 37435 -----ATACCTGACAGGATTCCGCGCGCGCGCGCGCGCGCGCTCT 37476
Qy 2096 ---LeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGlnSerGlnPro 2114
Db 37477 CCGCGCTCGCGCGCGCTCGCTCCGTGCGCGCGGTTCGCGCGGTTCGCCAACACACCCG 37536
Qy 2115 SerSerSerProLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThr 2134
Db 37537 CCGGCCCCACCG---TCGCGCCCGGTTCGCGAGGTGCGGTGGCGCGGTGCGGATCAGC 37593
Qy 2135 -----LeuAlaGlnHisIle----- 2139
Db 37594 GGTCCGCCCGGTGAGGGCTTGGGTGGGCTCGTTGATCGGCTTGAGGATTTGTTGTCGAGG 37653
Qy 2140 -----SerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAla 2157
Db 37654 GTGTGAGTGGGTGCTGCGCGGGCGGTGAATCGCTCTCGACCTAGTAGTGCCTGCGCT 37713
Qy 2158 ProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArg 2177
Db 37714 AAGCCCGCGCGCGCGCGCGTGC---GCGGTGCGCGCATGCTCCGCGCTACCATCCGCTTA 37770
Qy 2178 ArgPro---ProSerAspLeuTyrLeuProProProAspHisGlyAlaProAla----- 2194
Db 37771 CCGCATTCGCGATCAGCACGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 37830
Qy 2195 -----ArgGlySerProHisSerGluGlyGlyLysArg 2205
Db 37831 GCGTGGCGCGCGGTTCGCGCGGTTCGCGGTTCGCGACGCGCGCGCGCGCGCGCG 37890
Qy 2206 Ser-----ProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGly 2221
Db 37891 GCCCCACCGGTTCGCGCGCGCGCGCGAAG-----GACCGCGCG 37929
Qy 2222 IleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
Db 37930 CCGCGCGCGCGCACCGCGCGCGCGCGATGAGCAGACCGCGCTTCGCGCGCGCGCGCG 37989
Qy 2242 TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer 2261
Db 37990 GCCCGCGCGGTTCGAGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38031
Qy 2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
Db 38032 -----CCGCGC----- 38037
Qy 2282 MetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
Db 38038 -----GAG 38040
Qy 2302 ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
Db 38041 CCGAAC-----AACGTGCGCGC----- 38058
Qy 2322 ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
Db 38059 -----TCGCTCCGATCCACCGCGCACCGCGCGCGCGCGCGCGCGCGCGCG 38097
Qy 2342 LeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerPro 2361
Db 38098 TTGAT-----CCG 38106
Qy 2362 ProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMet 2381
Db 38107 CCGCTGCC--GCCG 38165

```
Qy 1058 -----ProProArgGluValIleLysAlaSerPro---Hi 1068
Db 34791 CGAGGTGCTCGCGGAGCATTAATCGCTAGACCG-----AGCAGCAGCCCGCTGAC 34844
Qy 1068 sAlaProAspProSerAlaPheSerTyAlaProProGlyHisProLeu---ProLeuG1 1087
Db 34845 GCCTCCCACTCCGCGCTTGCCTCGCGGCAATCCCACTCCGCTACCGCGCTTACCGCATTGCC 34904
Qy 1087 yLeuHisAspThrAlaArgProValLeuProArgProThrIleSerAsnProProPr 1107
Db 34905 GATCAACACGCGGTGCGCGCATCCGCGCTTGCCTCGCGGTACCGCGGTGGCGCCACC 34964
Qy 1107 oLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerG1 1127
Db 34965 GTTACCGCGGTGCGCGCTTACCG-----ATCAGCCC 34997
Qy 1127 nGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyPr 1147
Db 34998 GGG-----GTGCG-----CCAGCCCCACC 35018
Qy 1147 oValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa 1167
Db 35019 GATCCCGCGCGGAGCGCTCCCACTCCGCGCTTGCCTCGCGCGCGCC--GGAGCGCA 35077
Qy 1167 lLysGlnGluLeuSerProArgGlyGlnAla-----GlyProProGluSe 1183
Db 35078 AGACGTGCGCGCTTGCCTCC-----GGGCGCGCTTGCCTCGCGCGGAGCGCAATC 35134
Qy 1183 rLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerVa 1203
Db 35135 CGCGCGCGCGCGGAGCGCGGAGCGGAGCGGAGAGAGAGCGCCAGCGTTCGCGCGCGCGCG 35194
Qy 1203 lProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaI1 1223
Db 35195 C---GGCGCGCTATGCGCGCGCGCGGTGAGAGTACCGCGCTCCCGCGGATTCCGCGG 35251
Qy 1223 eThrTyArgGlySerIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrI1 1243
Db 35252 CGCGCGCGCGCGCG-----35267
Qy 1243 eThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLe 1263
Db 35268 -----GAGGCGGAGCATGCGCGCAT 35287
Qy 1263 uProLysGlyHisValIleTyGluGlyLysGlyHisValLeuSerTyArgGlyG1 1283
Db 35288 TGCGCGCGCGCGCGCTCCCGCGCGCGACAG-----GCTGTGTCGCGCG 35335
Qy 1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisG1 1303
Db 35336 TGCGCGCTTCCCGCTCGCGCGAA-----CA 35362
Qy 1303 uThrAlaAlaProLysArgThrTyAspMetMetGluGlyArgValGlyArgAlaIleSe 1323
Db 35363 GCGCGCGCGCGCGCGCGCGCGCTCGCGGAGCTGCTGTGCGGAGAACCGCGCAT 35422
Qy 1323 rSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHi 1343
Db 35423 GCGCGC-----GGTGGCGCGCGCGCGAGCGAGCGCGC--CAGCGCCACCGCGCCCA 35472
Qy 1343 sHisLeuLysGluGln-----HisHisIleArgGlySerIleThrGlnGlyIlePr 1360
Db 35473 CCGCGCGCGCGGAGCTGCGCGCGCGCGCGCGGATCGCGCA-----CC 35514
Qy 1360 oArgSerTyValGlnAlaGlnGluAspTyLeuArgArgGluAlaLysLeuLeuLysAr 1380
Db 35515 CCGCGGTGGGGAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTGCACTG 35574
Qy 1380 gGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyLysThrG1 1400
Db 35575 CCGTTC--GTGAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35604
Qy 1400 nAlaLeuGlyProLeuLysLeuLysProAlaHisGlyGlyLeuValAlaThrValLysG1 1420
Qy 1058 -----ProProArgGluValIleLysAlaSerPro---Hi 1068
Db 34791 CGAGGTGCTCGCGGAGCATTAATCGCTAGACCG-----AGCAGCAGCCCGCTGAC 34844
Qy 1068 sAlaProAspProSerAlaPheSerTyAlaProProGlyHisProLeu---ProLeuG1 1087
Db 34845 GCCTCCCACTCCGCGCTTGCCTCGCGGCAATCCCACTCCGCTACCGCGCTTACCGCATTGCC 34904
Qy 1087 yLeuHisAspThrAlaArgProValLeuProArgProThrIleSerAsnProProPr 1107
Db 34905 GATCAACACGCGGTGCGCGCATCCGCGCTTGCCTCGCGGTACCGCGGTGGCGCCACC 34964
Qy 1107 oLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerG1 1127
Db 34965 GTTACCGCGGTGCGCGCTTACCG-----ATCAGCCC 34997
Qy 1127 nGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyPr 1147
Db 34998 GGG-----GTGCG-----CCAGCCCCACC 35018
Qy 1147 oValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa 1167
Db 35019 GATCCCGCGCGGAGCGCTCCCACTCCGCGCTTGCCTCGCGCGCGCC--GGAGCGCA 35077
Qy 1167 lLysGlnGluLeuSerProArgGlyGlnAla-----GlyProProGluSe 1183
Db 35078 AGACGTGCGCGCTTGCCTCC-----GGGCGCGCTTGCCTCGCGCGGAGCGCAATC 35134
Qy 1183 rLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerVa 1203
Db 35135 CGCGCGCGCGCGGAGCGCGGAGCGGAGCGGAGAGAGAGCGCCAGCGTTCGCGCGCGCGCG 35194
Qy 1203 lProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaI1 1223
Db 35195 C---GGCGCGCTATGCGCGCGCGCGGTGAGAGTACCGCGCTCCCGCGGATTCCGCGG 35251
Qy 1223 eThrTyArgGlySerIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrI1 1243
Db 35252 CGCGCGCGCGCGCG-----35267
Qy 1243 eThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLe 1263
Db 35268 -----GAGGCGGAGCATGCGCGCAT 35287
Qy 1263 uProLysGlyHisValIleTyGluGlyLysGlyHisValLeuSerTyArgGlyG1 1283
Db 35288 TGCGCGCGCGCGCGCTCCCGCGCGCGACAG-----GCTGTGTCGCGCG 35335
Qy 1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisG1 1303
Db 35336 TGCGCGCTTCCCGCTCGCGCGAA-----CA 35362
Qy 1303 uThrAlaAlaProLysArgThrTyAspMetMetGluGlyArgValGlyArgAlaIleSe 1323
Db 35363 GCGCGCGCGCGCGCGCGCGCGCTCGCGGAGCTGCTGTGCGGAGAACCGCGCAT 35422
Qy 1323 rSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHi 1343
Db 35423 GCGCGC-----GGTGGCGCGCGCGCGAGCGAGCGCGC--CAGCGCCACCGCGCCCA 35472
Qy 1343 sHisLeuLysGluGln-----HisHisIleArgGlySerIleThrGlnGlyIlePr 1360
Db 35473 CCGCGCGCGCGGAGCTGCGCGCGCGCGCGCGGATCGCGCA-----CC 35514
Qy 1360 oArgSerTyValGlnAlaGlnGluAspTyLeuArgArgGluAlaLysLeuLeuLysAr 1380
Db 35515 CCGCGGTGGGGAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTGCACTG 35574
Qy 1380 gGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyLysThrG1 1400
Db 35575 CCGTTC--GTGAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35604
Qy 1400 nAlaLeuGlyProLeuLysLeuLysProAlaHisGlyGlyLeuValAlaThrValLysG1 1420
Qy 1420 uAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPr 1440
Db 35611 -GGCGGAAGAGAGGAGCGCGCTTGC-----CCAGCCCCCGCC 35648
Qy 1440 oLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyAr 1460
Db 35649 GCGCGCGCGCGC--CCGCGCGTGAGGGCTACTACGCGCGCGC-----35690
Qy 1460 pThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerPr 1480
Db 35691 -GGCGCGCGCGCGCGCGCGCGCGCGCAACAGCATGGCGTTGCCGCG-----GGCTCCGC 35743
Qy 1480 oGlyArgThrPheProProVal---HisProLeuAspValMetAlaAspAlaArgAlaLe 1499
Db 35744 CGACCCCGCGCATCCCACTGCTGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35803
Qy 1499 uGluArgAlaCyeTyArgGluSerLeuLysSerArgProGlyThrAlaSerSerG1 1519
Db 35804 TGAGCGCGC-----GGCGCGCGCGTTGCCGCGG 35833
Qy 1519 yGlySerIleAlaArgGlyAla-----1526
Db 35834 CGCGCGCGTTGCCGCGCGCGCGCGCGCGTGTGACGCGCGCGCGCGGATCTCTCGCGCGCGC 35893
Qy 1527 -----ProValIleValProGluLeuGlyLysPro-----1536
Db 35894 CGTTGCCGATTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGG 35953
Qy 1537 -----ArgGlnSerProLeuThrTyArgGluAspHi 1546
Db 35954 CGCGGTGCGCGATCAACGCGCGCGCGGTATTTCGCGAGAGAACTCGTTGA--TCGGATCC 36012
Qy 1546 sGlyAla-----ProPheAlaGlyHisLeuProArgGlySe 1558
Db 36013 AGCAGCGCGGACACCGCGCGCGCTCGCGCGCGCATAGCGCGCGCGCGCGCGCGCGCAAT 36072
Qy 1558 rProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLe 1578
Db 36073 GCTTCAGCAACTGGGCGATGAACAGCGCTGCGCTT-----GGGCGCTGAGCGCGCTGATAG 36126
Qy 1578 sAlaSerGlnAspArgLysLeuThr-----SerThrProArgGluIleAlaLy 1594
Db 36127 GCCTGGCGTGGGCGCGCAACAGCGCGCGATGCTGTCGACACCTCGT-----CG 36177
Qy 1594 sSerProHisSerThrValProGluHisHisProHisProIleSerProTyArgGluHisLe 1614
Db 36178 GCGCGCGCGCGCATCAGTGGCGGTGGCGCGC-----36214
Qy 1614 uLeuArgGlyValSerGlyValAspLeuTyArgSerHisIleProLeuAlaPheAspPr 1634
Db 36215 -----CGGCTGCGTTGCGCGCTGATGCTCGAT 36243
Qy 1634 oThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyTyLeuProArg-- 1653
Db 36244 CCGAGACTGGCCAAATCCGTTGCG-----CTGCGCGCGATA 36279
Qy 1654 -HisLeuAlaPro-----AsnProThrTyProHisLeuTyProProTy 1668
Db 36280 ACCTCTGCGCGCGCAATCAAAACGACATCTGCACACCTCCCAATACGATGACCG-----36334
Qy 1668 rLeuIleArgGlyTyProAspThrAlaAlaLeuGluAsnArg-----1682
Db 36335 -CTCTGTCATGCCGACCGCGGGAACGTCACAGCAAAAATCGGAGTAAGAAGCATCCCA 36393
Qy 1683 -----GlnThrIleLeuAsnAspTyIleThrSe 1692
Db 36394 TTTCACGAGCAACACCTGGGGGGTTTGGTCAAACTCTGTAAGCACTTCGTG-----36448
Qy 1692 rGlnGlnMetHisHisThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgG1 1712
```

32844 GTCGAGAGTTTTCGACAGCAGCAGGTTCTCGGGCCGGGGTCTTCGGAAGGTGATG 32903
Db
465 -----GluCysValLeuTyrTyrTyrLeuThrLys 474
Qy
32904 GGGCTGGCGGTTGACGGCTTCACGAATGCGAATGTCATCGCTCGTGTGACGAATCTC 32963
Db
475 LysAsnGluAenTyrLys-----SerLeuValArgArgSerTyrArgArg- 490
Qy
32964 GAAGATGCGTTCGATTCGTTAGCCGAGAGAACAGCCACACTAGGTTCGGCACTGT 33023
Db
491 ---GlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Qy
33024 GAAGGGTCTGCGCGGACAGCTGATCAACAGGGGGCGGCGGTTCGGTGAATTCG 33083
Db
510 GlnProMetProArgSerSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 527
Qy
33084 ATGGTCACCGCGGTGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 33143
Db
528 -----GluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAsp 544
Qy
33144 GTGCTGTCCAGGCTCACTTCGCGAGTTCACTTCAGACCCCTCGTCGACACTTCGGAC 33203
Db
545 LeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaVal 564
Qy
33204 AACACCGTATTCAGGTTAGTTCGATACCTGGCGCGGACCGGTCCGGCGTCAAAAACG 33263
Db
565 AlaSerLysGlyArg-----LysThrAlaAsnSerGlnGlyArgArgLysGly 580
Qy
33264 GGGACGGTTGGCGGGCGCGCGGTACGGCGCGCGGAGCTCCCGCGGAGGCGGTCT 33323
Db
581 ArgIleThr-----ArgSerMetAlaAsnGluAla 590
Qy
33324 TCAGTTCACAGCGCCCGAGGATTAACACGCTCGCGGGCGCGGCTTTCGGGTAGCCAGT 33383
Db
591 AsnSerGluGlu----- 594
Qy
33384 AGTTTCACGATTGACACGGTGTGGCGATGATGCTCAGGATAGCGCATCTACGGCAT 33443
Db
595 -----AlaIle-ThrProGlnGlnSerAlaGluLeuAlaSerMet 607
Qy
33444 CATCTCGCGTGAGCAACTGATCCCTCAACGCGGTGTGTGTCGAGGTCTGCTGTCTAT 33503
Db
607 tGluLeuAsnGluSerSerArgTyrThrGluGluGluMetGluThrAlaLysLysGlyLe 627
Qy
33504 G-----GCAAGCCGTGTAGTCCGTCTCGCCAGCAGCAGCGCCGCG- 33546
Db
627 uLeuGluHisGlyArgAsnTyrSerAlaIleAlaArgMetValGlySerLysThrValse 647
Qy
33547 -TGCGCCCGCACCGGGCGTGGCCCGGCTTTCGCG----- 33582
Db
647 rGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuG 667
Qy
33583 -CGTTGCGCG-----CGTTGCGCGGTTCGCGATCAGCACGGCGGTTCGCGCG 33629
Db
667 nGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaPr 687
Qy
33630 ACACACCGCTGCGCGCGGTACCGG---CGCCAAACCCCGCGCAACCCCGCTCACCGCC 33686
Db
687 oAlaAlaAlaSerGluGluAlaAlaPheProValValGluAspGluMetGluAl 707
Qy
33687 G-----TTGCCGAACACCCCGCGTGGCCACCG----- 33714
Db
707 aserGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSe 727
Qy
33715 -----TCACCGCGGTGCGCGGTACCGCGGCTTACCGCGCTTAGAGCGTT 33752
Db
727 rGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAs 747
Qy
33753 GGCACCGCTGCCCG-----GGCGCGCGCGCGCGCGAGCGAAGAGCAA 33800
Db
747 pThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyPr 767
Qy
33801 GCGCGCGTTCCCGCGCGCGCG-----CC 33824
Db

767 oLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrProPro-- 786
Db
33825 GGGCGCGCTTGTGTGATGCTGTAGTGTCTGCGCGCGGTGCGCGCGCGCGCGCG 33884
Qy
787 -----ArgArgThrSerArgAlaProLeuGluProThrProAlaSerGluAlaThrG 804
Db
33885 GCGCGGAAGCCGAGAGTAAGCGCGTTCGCGCGGTTCGCGCGCGCGCGCGCGCAAG 33944
Qy
804 y-----AlaProThrProProAlaProProSer----- 814
Db
33945 GGAGCTGGCGCCACCGCTCCCGCGCGCGCACCGAGGCGCGCGAGGAGTAGGCGCG 34004
Qy
815 -----ProSerAlaProProValValProLysGluGluLysGluGluThrAl 832
Db
34005 GTTCCCGCGGTGCGCGCGCGCGCGGTGTGTGATCCCGGACCTCCGAG-----CC 34055
Qy
832 aAlaAlaProProValGluGluGluGluGluGluLys-----ProProAl 847
Db
34056 GCGCGCGCGCGGTGCC-GCCGCGCTCCGAACAGTCCGCGTTCGCGCGCTCCACCGG 34114
Qy
847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGlu-----Pr 861
Db
34115 CCGCGAAGTTGTCGCGCGCGCGCGCGCGGTGCGCGAGTTCGAACAGTCCGCGTTCGCG 34174
Qy
861 oValLysSerGluCysThrGluGluAla---GluGluGlyProAlaLysGlyLysAsp-- 879
Db
34175 CGTTCGCGCGGTGCGTGAACCGCGCGCGCGCGCGGTGCGCGTTCGCGAAGTCC 34234
Qy
880 -----AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAl 893
Db
34235 CGCGGTTCG 34294
Qy
893 aGluLysLysGluGlyGlySerGlyArgAlaThr-----ThrAlaLysSerSe 909
Db
34295 C-----GTCG 34339
Qy
909 rGlyAlaProGlnAspSerAspSer-----AlaThrCysSerAlaAspGluVa 926
Db
34340 CG 34394
Qy
926 lAspGluAlaGluGlyLysAsnArgLysLeuSerProArgProSerLeuLeuTh 946
Db
34395 -----GCG 34415
Qy
946 rProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLysGlnLe 966
Db
34416 -----GGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 34467
Qy
966 uLysGlnArgAlaAlaAlaProIleProIleGlnValThrLysValHisGluProProAr 986
Db
34468 ATCCCGCGCGTCCCG-CCATTGCCACCGGTACCG---AACAAAGATC-----CCGCGCGC 34517
Qy
986 gLusAspAlaAlaPro-----ThrLysProAlaProProAlaPr 999
Db
34518 CCG 34577
Qy
999 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGly-----Se 1017
Db
34578 GCGGTTCGGAACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 34637
Qy
1017 rSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAl 1037
Db
34638 GCGCGCGTTCGGAACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 34674
Qy
1037 aGluAlaGlnLysLeuProGlyAspPro-----ProCysTrpThrSerGlyLeuProPh 1055
Db
34675 -----TCACCGCGGTTCGAGGAGTCCCGTTCGCGCGCGCGCGCGCGCG 34730
Qy
1055 eProVal----- 1057
Db
34731 GAGCGTCTGGAAGGCGTCTTCACCATTCAGACACATTTTGTCTGACAGGTGTGCGAGTGG 34790

QY 2051 GluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeu 2070
 DB 64745 GCAGGCAACACCGCGTCCGCGC-ATCAGCCGCGAGCGGCTTCAAGTCGAGTGGACCGG 64687
 QY 2071 --ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyLeuLeuArgProLys 2089
 DB 64686 AACCAAGAGCATTCGCCCGGCAACGCAACGCGCATCTCGAA- 64645
 QY 2090 GlnProGlyProValLysLeuGly- 2097
 DB 64644 CAGCGCAACCGCTCCGCGCGGCAATCGCGGAGTGTGCGGCAATGCGGTACGCGCC 64585
 QY 2098 -----GlyGluAlaHisLeuProHisLeuArgProLeuProGluSer- 2112
 DB 64584 CGCCGCATCGAGCGCTCCGCTCATCCCGCTGGCTGCTCCACAGGCCCAAGCAGCGA 64525
 QY 2113 -----GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHis 2129
 DB 64524 CACCGCAGGACCGCGCGTCCGCGGTGTGAGCCAGGCGCTCCAGGAACGCGTTCGC 64465
 QY 2130 GlnArgValValThrLeu- 2135
 DB 64464 CGCCCGTAGTTTCCCTGACCCCGCAGCCCCCGCCAGCGCGGAGACGAGAGAC 64405
 QY 2135 ----- 2135
 DB 64404 GAAGAACGAGAGATCCAGATCAAGAGTCAACTCGTGCAGGTGCCACGCGCGCGCTT 64345
 QY 2136 ---AlaGlnHisLeuSerGluValThrGlnAspTyrThrArgHisProGlnGln 2154
 DB 64344 CGGCGCGCAACACTTTGCCACGTGTCTGT-GTGTACGAGCTCTGACGACCCGCGTGGCA 64286
 QY 2155 LeuSerAlaProLeuProAlaProLeu- 2165
 DB 64285 GTGCGCGCGCG- -CATGCACCACTCCGCGCAGCGGGAACCTCTCCGAGATCGTCA 64229
 QY 2166 PheProGlyAlaSerCysProValLeuAspLeuArgProProSerAspLeuTyrLeu 2185
 DB 64228 GCAGCGGTCCCACTGGCGCGCATCCCGACATCGCAGCTCCGACCTCCGACCGAGCTC 64169
 QY 2186 Pro- -ProProAspHis- 2197
 DB 64168 CCAGGCGCCCAACTATCCACAGGTCTGTGACCCGCGGCGCTCCAGCGCGTTCGCC 64109
 QY 2198 ProHisSerGlu- -GlyGlyLysArgSerProGluProAsn- 2210
 DB 64108 CCGGAGCACCAGCGCGCGGACACCGCGTCCGCGCAACCAATGCCGGGCAACCAACCGCG 64049
 QY 2211 -----LysThrSer- 2217
 DB 64048 CCAGCAGCGCGTACCGCGGAATCAACACCTCCATCCGACCGGGCACTGGATCC 63989
 QY 2218 GlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSer 2237
 DB 63988 GGCTGTCTCGGCGGCAACATCGCGCCAGGCGAGGCA- -CCCGCACTTCAC 63938
 QY 2238 ArgSerAlaValTyrProLeuLeuTyrArgAsp- 2248
 DB 63937 CGGCGCGCAAGCGCAGCTGCGGTCTCCGGAACGACGCGCGCGGCGAGCGGACGCAATG 63878
 QY 2249 -----GlyGluGlnThrGluPro 2254
 DB 63877 ATTCGGGGTTCGCTCCACATCGACCAACAGCAATCGTCCGGGTTCTCCGACTCGCGCG 63818
 QY 2255 Ser- 2259
 DB 63817 AACGCAACAAACCCAGCAGCAGCATGCAACAGGTCCGGAACACCTCCGCGGAGCTG 63758
 QY 2260 LysSerProGlyAsnThrSerGlnProAlaPhePheSerLysLeuThrGlu- 2277
 DB 63757 TCGAGCGCGCACCACTGCAACCAACCAAGGAGATCCGCGCAATCCGCGCGCGCGCA 63698

QY 2278 -----SerAsnSerAlaMetValLysSerLys- 2290
 DB 63697 ACCAATCTCGATCAGGTCAAGCATCGTGTGTGTCGCGCAGCTGCTCCGGAACCACT 63638
 QY 2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
 DB 63637 CCAAGTCTCGCGAGCAGCAGCGCATGTCACCAACCA- -CGTCCGCGG 63590
 QY 2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln 2330
 DB 63589 GCG- -GGTCAGCAACCCCGTCAGTCCGGGTGG-AGTTCGTCCCAAGTGC 63540
 QY 2331 AlaValGlnGluHisAlaSerThrAsnMet- -GlyLeuGlu 2343
 DB 63539 GAGCCCGAGCTCGCGGAGATTCTGTCATTCGCGATCACTGCCAGGAGACCGGCTTGCA 63480
 QY 2344 Ala- - 2347
 DB 63479 GCGCGCGGCGAGATTCTCCAGTCCGACCCGGAACAGGACTGCTGTTCATTCGACGG 63420
 QY 2348 LysAlaLeuMetGlyLysTyrAspGlnTyrGluLys- - 2360
 DB 63419 GCGCTGCCAATCTGTCGCGCAGCAGCAGCAATCGATCGAGAGCAGCAG 63360
 QY 2361 -----ProProLeuSerAla- -AsnAlaPheAsnProLeuAsnAlaSer 2374
 DB 63359 GCTTGGCAGTCTGCTGCTGTAACACGACACCGCATTCGCGCGCCGGAACGAGG 63300
 QY 2375 AlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu 2394
 DB 63299 TGCACCCGAAAGTGGCGCGCTTTCGAGGACACACTTGCACGCGAAGCGG 63240
 QY 2395 ThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
 DB 63239 AGCATCTCTGCTCGGATCTCT- -TCGCTGCGCCGAGCGCTCGAGTGC 63189
 QY 2415 LysSer- -ProAlaProGlyLeuAlaSerGlyAsp 2425
 DB 63188 GCATCCAGCAGCAGCGGTGGATTCCGAACCTCCGCGCTCTTCGCGGTGTGGTGGG 63129
 QY 2426 ArgProProSer- -ValSerSerValHisSerGlyGlyAspCysAsnArgArgThr 2443
 DB 63128 AGCGCGACCTCGGCGAAGATCTGCTCTCCGCTTCCAGCGCGCTCAAC- - 63078
 QY 2444 ProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThrProPhePro 2463
 DB 63077 -----CGCTGGAACGCGCGCGGCGGAATCGAAGCCCGCTCAGTCCGCGG 63033
 QY 2464 TyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGly 2483
 DB 63032 GCGTA-GAAGCTGTCAGCGGAACCGGACAGCATCTGTGGCGCGCAATCCCATCGG 62974
 QY 2484 LeuProAlaGly- -SerGlyProLeuAlaGlyPro 2494
 DB 62973 GCCCTGGCGCGGCTCTGCTGATCGTGTGAGGTCAACGTGCGGTGCGTGGCAGGTCCA 62914

RESULT 64
 AAI99682_03
 Continuation (4 of 45) of AAI99682 from base 300001 (Mycobacterium tuberculosis strain H
 Wp Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
 Wp Fragment Name Begin End
 WP AAI99682_00 1 110000
 WP AAI99682_01 100001 210000
 WP AAI99682_02 200001 310000
 WP AAI99682_03 300001 410000
 WP AAI99682_04 400001 510000
 WP AAI99682_05 500001 610000
 WP AAI99682_06 600001 710000
 WP AAI99682_07 700001 810000
 WP AAI99682_08 800001 910000
 WP AAI99682_09 900001 1010000
 WP AAI99682_10 1000001 1110000
 WP AAI99682_11 1100001 1210000

Db 66603 CGCATCCAGCGTGGAAATATATCCGACGCTACCCAAATAGGACAAATCGCTCGCGCTG 66544
Qy 1429 ProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlu 1448
Db 66543 AGCCTGAGCGCTCCGGTGTCTCCCGACACATCAACGGGACCGCTGGAAGTCCAC 66484
Qy 1449 GlySerIleThrGlnGlyThrProLeuLysTyrAsp 1460
Db 66483 GGTGATTCCGAATCGGGACCGCTCTTGTCCGAGTCTCGTCTGCTGCTGTTTCGAGGCTG 66424
Qy 1461 -----ThrGlyAlaSerThrThrGlySerLys---LysHisAspVal 1473
Db 66423 TTCAGGATGACGTGTGCTGTTGGTGGCTGATCCCGAAACAGCACGCCACCCCGCG 66364
Qy 1474 ArgSerLeuIleGlySerProGlyArgThrPhe-----ProPro 1486
Db 66363 AGGATGGCTGTTCGGGGCCAGGGGTGTTCT-CCGTGAGGAGCTGGACATCCCGCCCGCG 66305
Qy 1487 ValHisProLeuAspVal-----MetAlaAspAlaArgAlaLeuGluArgAla 1502
Db 66304 ACCAATCCACTTCGAGCTAGTTTCATCCACATGCAACGTCCGCGGCGAGCTGCCGCTGCC 66245
Qy 1503 -----CysTyrGluGlu 1506
Db 66244 GCATCGCCATCACCATCTTGATCACACGAGCGACCCGCGCGCTGCGTATGACCAA 66185
Qy 1507 ---SerLeuLysSerArgProGlyThrAlaSerSerGlyGlySerIleAlaArgGly 1525
Db 66184 TATTCGACTTCACGACCCCAACACACAGCCGATGCTATCCCGCCCTGCGCGT--- 66128
Qy 1526 AlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAsp 1545
Db 66127 -----AAGTGGCAATCAGAGCCTCGCGCTCAATCGAT 66095
Qy 1546 HisGlyAlaProPheAlaGlyHisLeu-----ProArgGlySerProValThr 1561
Db 66094 CACCAAGCCTGGTTCAGTCCCATCGCGCTCCACGCGGTCCACATCAGACACCGCACAAAC 66035
Qy 1562 MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGln 1591
Db 66034 CAGCACTCGCCAAACGCT-----CGGTGATCACCCTGCTGCGAGGCCCATTCG 65984
Qy 1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db 65983 -----GGCGCGTCAAAACCATTCGACG----- 65963
Qy 1602 GluHisHisPro-----HisProIleSerProTyrGluHisLeuLeuArgGly 1617
Db 65962 ---CACCATCTGATTACCGCGCTACCCCGACGCGCCAGAACCGCGTGCCTGCGTGC 65906
Qy 1618 ValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIle 1637
Db 65905 CCGCGG-----CATCCGACAGCCCTCCAAACACACC-AGCCCGCGCCCTTCT 65859
Qy 1638 ProArgGlyIleProLeuAspAlaAlaIleTyrLeuProArgHisLeuAlaPro 1657
Db 65858 CCCCACCGGTGCCA---TCGCGAGCGCGCAACAGCTTGCACCGGCCCATCCGCGCC 65802
Qy 1658 AsnPro-----ThrTyrPro-----HisLeuTyrPro 1666
Db 65801 AAGCCCGCTTGACGCGAAACATCCACGAAATCCCGCGCTGCGCATCACCGTCACACCA 65742
Qy 1667 ProTyrLeuIleArgGlyTyr---ProAspThrAlaAlaLeuGluAsnArgGlnThrIle 1685
Db 65741 CCGCAAGAGCAGCATCATCTACAGCCGAGTGTTCACCGCGCAGTGCAGCGCC 65682
Qy 1686 IleAsnAspTyrIleThrSerGlnGlnMetHisAsnThrAlaThrAlaMetAlaGln 1705
Db 65681 ACCAACGACGA-CGAACA---CGCGTATCCACCTCATCCGCGCAGGACCTCGAAACCAA 65626
Qy 1706 ArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAla 1725

Db 65625 CGA-----ATACGCAACCCGACCCCGAAACACACCCCTCCCGCGCTGC-CATTACCTA 65576
Qy 1726 AlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValPro 1745
Db 65575 GATAACCTTCGAAC-----CCT 65558
Qy 1746 ProThrPro----- 1748
Db 65557 CCGCGCGCTCGTATGAACCCGCGCGCTAGTGTGTGTCATCAGCCCGCGCAACACGC 65498
Qy 1749 -----GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln 1766
Db 65497 CGGTCCGCTCCCCCGCAACGAAACGCGATCAATCCCGCGCGCTCCACGGTTCCTCAAG 65438
Qy 1767 ProPheSerSerArgHisSerSerProLeuSerPro-----GlyGlyProThr 1783
Db 65437 CGACTCCAGCAGCAACCGCTGCTCGGATCCATCGCCACGCTCAGCGGGGAATCC 65378
Qy 1784 -----HisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAsp 1799
Db 65377 CGAAGAAACCGGATCGAATCTCTCGCATCTCGCAGGAAGCCACCCCTCACACATACG 65318
Qy 1800 ArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThr 1819
Db 65317 AGGTCCCAAGAC---GATCGGATCGGATCGA-----ACA 65285
Qy 1820 ThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db 65284 ACCCGCTGATCCAGCCACGATCTCCGGAACTCCGCCACCGCATCCAAAGCCCGCG 65225
Qy 1840 SerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln 1859
Db 65224 CCACCA-----AACGCCAAAGCTCTCCGCGAGGACACCCACCCCGGATAAC 65177
Qy 1860 HisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHis 1879
Db 65176 GACAGCCCA---TCCCGACGATCACAACGGGATCGTCAGCAGTCACCG----- 65132
Qy 1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
Db 65131 ---CCGCCACGCGCGGTGCGGAGTCGGAGCTCCACGCCACCAATCAGTTGTGCA 65075
Qy 1900 SerThrSerThrSerSerProValArgProAlaAlaThrPhePro-----ProAlaThr 1917
Db 65074 GCAAGAACCAACCAACCGCGGTGTGCGATAGTCGAACACCGCAGTAGCCGCAACC 65015
Qy 1918 HisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu 1937
Db 65014 GAACGCCCA----- 65006
Qy 1938 LeuProLysGluAlaProArgValAlaArgProGluArgProAlaAspThrGlyHis 1957
Db 65005 -----GCACCCCGCAAGCGGTTCGAAACTCCACACCGCAGCAACGAATCAA 64958
Qy 1958 AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLys 1977
Db 64957 -----AACCCAAATCCCGAAAG 64940
Qy 1978 GlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThr 1997
Db 64939 CAGATCGGCGCGGACCTCGACCGCACCCG-----AATGCC 64904
Qy 1998 ProAlaLysAsnLeuAlaProHisHisAla---SerProAspPro----- 2011
Db 64903 CCAGCA-----CCGCGCAACATGCTCGCGCACCAACCCACACGCTGGTTCCT 64853
Qy 2012 ---ProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
Db 64852 GATCCAGTCCGCGAGCTCCGAGCGCA-CCAACCAAGACCGTTGACGTCGCGGACTCG 64794
Qy 2031 PheSerIleGlnLeuLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
Db 64793 TTGACC-----GGCTCCCGCGCACGCGCGCTGGCACACGAGTCCGCAACAAAC 64746

```
QY 858 aGluProValIysSerGluCysThr---GluGluAlaGluGluGlyProAlaIysG1 877
Db 68562 CGCGTCGCGAGCGCGAGCGCTCGAGCGCGCATCCGCAACAGCTCGGTCGAGAGGTA 68503
QY 877 yLysAspAlaGluAlaGluAlaThrAlaGluGluGluAlaGluAlaGluAlaGlu 897
Db 68502 TCCGAGCGCC---ATCGCATCGGCTCGCGCGCGCGGCGGATTTCCGCGAGAACCTCGTC 68446
QY 897 uGlyGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGlnAsp----- 914
Db 68445 ACCAGCGCCATCGCGCTGAGCGCTGGAACACAGCGCGCTACTCGAACCGTTTCC 68386
QY 915 -----SerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
Db 68385 GCGCAGCGGACGTAGAACTCATCGAGAGCAATCGACTCGGCACCTTCAGCGGACCAATC 68326
QY 929 a-Glu-----glylyAspL 934
Db 68325 AGGAACCGCCCTACACCGGTGTGCGCAGCAGTTGCGCGGTGGGTGGCGCTCCA 68266
QY 934 yAsnArgLeuLeuSerProArgPro-----SerLeuLeuThrProThrGlyAsp----- 950
Db 68265 ACCCCACCTCTCGACCCACCGCGCGCGGGAATACAGCAGACACCTTCGCGCAGCGCC 68206
QY 951 -----ProArgAlaAsnAlas 956
Db 68205 TTCTGCGGTGCTTACCAGCAGCTGGATTTCCAGTGGCGTGTTCGTCGCCGACAGTTC 68146
QY 956 erProGlnIysProLeuAspLeuIysGln----- 968
Db 68145 TTCCCGAGTTTCAATTGAACCTGTGCTCCCTCCCGCAGCGCCACGACTACCGGACCGCG 68086
QY 968 ----- 968
Db 68085 CGCGGTTCGGGACACACCGCGGTGTTTCCAGGTGAGCTTCTCCACCCGCTCACAGCC 68026
QY 969 -----ArgAlaAlaAlaIleProProIleGlnValThrIysValHisGluP 984
Db 68025 GAGACGCGCCCTCGTGGCGGCGCCATCTCCAGGATGGCGGTGGCGGAGCAGCAGCTC 67966
QY 984 roProArgGluAspAlaPro-----ThrLysProAlaPro-----A 998
Db 67965 GCGCAGCGCGATGCTGGCGCAGCGCGATGCTGCTCCACCGCAACCGCGCGGTGAG 67906
QY 998 laProPro-----ProProGlnAsnLeuGlnProGluSerAspAlaProGlnG 1014
Db 67905 CACCATCCCATCACCATCGCGCGCAGCGCAACCGCGCGCCCGCGCGATGCTCCGCGCG 67846
QY 1014 laProGlySerProArgGlyLysSerArgSerProAlaProAlaAspLysGluA 1034
Db 67845 AGCAGCGCGATACCGTCACATCGC---CGTAATCCGATGTTTCCAGCGAGTACCGCTG 67789
QY 1034 laPheAlaAlaGluAla-GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
Db 67788 CCGCTGGAAGGCATACGTCGCAACTC-----GACCCCGCTGC----- 67750
QY 1054 ProPhePro-ValProProArgGluValIleLysAlaSerProHisAla-ProAspProS 1073
Db 67749 -----CCGCGTCCG-----CGAAATACGCGGACCATTCGACCGGAAACCC 67708
QY 1073 erAlaPheSerTyAlaProProGlyHisPro-----LeuProLeuGlyLeuHisA 1090
Db 67707 ACCGCTGTGCGCGCGCGCACCGCGCGACACAGTTTCCGTTTCCGTACGCTTTTACG 67648
QY 1090 spThrAlaArgProValLeuProArgProProThrIleSerAnProPro----- 1106
Db 67647 CAGCATCGGAATCGGATGCTGATCAGCCCAAGCAATGACATCGGGAATCAGGCGCGA 67588
QY 1107 -----ProLeuIleSerSerAla----- 1112
Db 67587 CAGCGTCGATAGGCGCGGATTCGAAGATCGTGGCGGACCCCTTGTGCGCGAGCACCG 67528
```

```
QY 1112 ----- 1112
Db 67527 GAGCGCTCGGAGCGAGCGGCTCTCGAACCTGACGACCCCAATACTCCGGGTAGC 67468
QY 1113 -----LysHisProSerValLeuGluArgGlnIleGlyAlaIles 1126
Db 67467 CATCACACCTGGGTGTCGAGCTCACCCGTCAACGTTCGACGACGCGGTAGT----- 67415
QY 1126 erGlnGlyMetSerValGlnLeuHisValProTy-SerGluHisAlaLysAlaPro----- 1144
Db 67414 -----CGCCACGCGGTAGTCCACCGCT 67393
QY 1145 -----ValGlyProValThrMetGlyLeuProLeuProMetAspProLysL 1160
Db 67392 GCGTGCATTTCCGCGCACTCTCTGCAACATCGGT-----CATCCGATACGAATGAACGC 67336
QY 1160 yLeuAlaProPheSerGlyValLysGlnGlnLeuSerPro-----ArgGlyG 1177
Db 67335 ATCGAAACCGCAACCGCGCACCGAATCCTTCCCATCCAGCGCGCGCGGTATGTC 67276
QY 1177 lnaAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgG 1197
Db 67275 GTCAAGCACTTCCGATCGCGGAGAGCACCACCGACTCAGGAG-----CGTT 67228
QY 1197 lyThrAlaLeuGlySerValProGlySerIleThrLysGlyIleProSerThrArgV 1217
Db 67227 G-ACCGCG-----GCGATACCGCGCACCATCA---ACCGATCGGCAACACGCGCGCA 67178
QY 1217 alProSerAspSerAlaIleThrTyr-----ArgGlySerIle 1230
Db 67177 GCTGCTCTCTCACGAGCAGCCCGCAACATGTCACCGCGCAGAGCAAGCAAGCGCTGCATCA 67118
QY 1230 hrHisGlyThrProAlaAspValLeuTyrlsGlyThrIleThrArgIleGleGluA 1250
Db 67117 ACCGGGACGACCGCG-----ACCATCCGAGCGCATCCGCGCAACG 67076
QY 1250 spSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIle 1270
Db 67075 ACAACAC-TCCCGCGCAGCACCGCGCGGAACT-CGCGACCGAATGGCCGACGACCAC 67018
QY 1270 yrGluGlyLysLysGlyHisValLeuSerTyrlsGlyGlyMetSerValThrGlnCys- 1289
Db 67017 ACCG-GGCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66968
QY 1290 -----SerLysGluAspGly-----ArgSerSerSerGlyPro-----ProH 1302
Db 66967 ACGGCAACAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66908
QY 1302 isGluThrAlaAlaProLysArgThrTyrlsAspMetMetGluGlyArgValGlyArgAlaI 1322
Db 66907 ACACCATCCCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66848
QY 1322 leSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGlu----- 1338
Db 66847 CGTCAACGCTCAGC-GAACACCGGAAAAACCGAGTAGAGCGCGCGCGCGCGCGCGCGCG 66789
QY 1339 -----ArgHisSerProHisHisLeuLysGlu-----GlnH 1349
Db 66788 CACTGTCCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66729
QY 1349 isHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrlsValGluAlaGlnGluA 1369
Db 66728 ACCAC-----CCAGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66696
QY 1369 spTyLeuArgArgGluAlaLysLeuLysArgGluGlyThrProProProProp 1389
Db 66695 GCTTCCACCGCGCAACCAACGTTTTCGGGTGCGGTCCCA-GCACCACCGCGCGCGCTC 66637
QY 1389 roSerArgAspLeuThrGluAlaTyrlsThrGlnAlaLeuGlyProLeuLysLeuLysP 1409
Db 66636 CAACCGCGCGCGCGGAGGAGCAACGAAACG-----C 66604
QY 1409 ro-AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIle 1428
```



```
Db 21758 CTGTCGTGGAAGG-----CATACGTTGGCAACTCGAGCTGCTACGCCGGGTACCGGC 21705
Qy 2403 lAlYsValSerGlyArGProSerSerArgLYsAlAlYsSerProAlaProGlyLeuAlas 2423
Db 21704 GAAAAACGCTGCCAGTCCACCTCACCACACCGGTGTGGAGATGCCCGAGGGCTGTGCAT 21645
Qy 2423 er-----GlyAspArgProSerValSerSerValHisSerGluGlyAspCysAsnA 2441
Db 21644 CACCGAAGCGGTCTCGTCCCGCTCCCTCGCATCAGTGGGATCG-----21601
Qy 2441 xGArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrp 2461
Db 21600 -----ACGAATCCGCCAACCCGATCGGATCTCGGTGC-----CACACTC 21561
Qy 2461 roPheProTyAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro---- 2479
Db 21560 CCGNACCAGTCCGCCCA-----CGGTTCGTCGCGGCCGAGCTCGACGACCGTGCC 21510
Qy 2480 -----ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProH 2495
Db 21509 TATGCCCTGGCGTCCAGGGCTCGACACCGTCCGCGAAACGAGACGGGTTTCACGCGCCT- 21451
Qy 2495 iHisAlaTrpAspGluGlyProLysProLeuLeuCysSer 2508
Db 21450 -----GGCGCACCCCAATACCCCTGTGTGTCTCA 21424

RESULT 63
AAZ21501/c
ID AAZ21501 standard; DNA; 80161 BP.
AC AAZ21501;
XX
XX 01-DEC-1999 (first entry)
XX
XX DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.
XX
XX Spinosyn biosynthetic enzyme; open reading frame; ORF;
KW insecticidal microides; arachnid; nematode; insect; polyketide;
KW polyketide synthase; PKS; extender module; initiator module;
KW acyl transferase domain; AT; acyl carrier protein; ACP;
KW beta-ketosynthase domain; KS; KR; dehydratase domain; DH;
KW enoyl reductase domain; ER; beta-ketoreductase; insecticide; ss.
XX
XX Saccharopolyspora spinosa.
XX
XX Key Location/Qualifiers
PH complement(1135..1971)
FT /*tag= a
FT /product= "ORFL16"
FT /note= "Protein involved in transcription control"
FT CDS 2024..2791
FT /*tag= b
FT /product= "ORFL15"
FT /note= "Keto acyl reductase"
FT complement(3416..4165)
FT /*tag= c
FT /product= "spnS"
FT /note= "Spinosyn biosynthesis protein S"
FT complement(4168..5325)
FT /*tag= d
FT /product= "spnR"
FT /note= "Spinosyn biosynthesis protein R"
FT complement(5363..6751)
FT /*tag= e
FT /product= "spnQ"
FT /note= "Spinosyn biosynthesis protein Q"
FT 7083..8450
FT /*tag= f
FT /product= "spnP"
FT /note= "Spinosyn biosynthesis protein P"
FT 8967..10427
FT /*tag= g
FT /product= "spnO"
```

```
FT CDS
FT /note= "Spinosyn biosynthesis protein O"
FT 10436..11434
FT /*tag= h
FT /product= "spnN"
FT /note= "Spinosyn biosynthesis protein N"
FT complement(11530..12492)
FT /*tag= i
FT /product= "spnM"
FT /note= "Spinosyn biosynthesis protein M"
FT complement(12696..13547)
FT /*tag= j
FT /product= "spnL"
FT /note= "Spinosyn biosynthesis protein L"
FT complement(13592..14785)
FT /*tag= k
FT /product= "spnK"
FT /note= "Spinosyn biosynthesis protein K"
FT complement(14799..16418)
FT /*tag= l
FT /product= "spnJ"
FT /note= "Spinosyn biosynthesis protein J"
FT 16556..17743
FT /*tag= m
FT /product= "spnI"
FT /note= "Spinosyn biosynthesis protein I"
FT complement(17749..18501)
FT /*tag= n
FT /product= "spnH"
FT /note= "Spinosyn biosynthesis protein H"
FT complement(18541..19713)
FT /*tag= o
FT /product= "spnG"
FT /note= "Spinosyn biosynthesis protein G"
FT 20168..20995
FT /*tag= p
FT /product= "spnF"
FT /note= "Spinosyn biosynthesis protein F"
FT 21111..28898
FT /*tag= q
FT /product= "spnA"
FT /note= "Spinosyn biosynthesis protein A"
FT /transl_except= (Pos:26940..26942, aa:Pro)
FT 21126..22379
FT /*tag= r
FT /note= "Beta-ketosynthase domain (KS1): part of the
FT initiator module"
FT 22692..23669
FT /*tag= s
FT /note= "Acyl transferase domain (AT1): part of the
FT initiator module"
FT 23793..24041
FT /*tag= t
FT /note= "Acyl carrier protein domain (ACPi): part of the
FT initiator module"
FT 24102..25349
FT /*tag= u
FT /note= "Beta-ketosynthase domain (KS1): part of extender
FT module 1"
FT 25683..26684
FT /*tag= v
FT /note= "Acyl transferase domain (AT1): part of extender
FT module 1"
FT 27582..28121
FT /*tag= w
FT /note= "Beta-ketoreductase domain (KR1): part of extender
FT module 1"
FT 28404..28649
FT /*tag= x
FT /note= "Acyl carrier protein domain (ACPi): part of
FT extender module 1"
FT 29024..30295
FT /*tag= z
FT /note= "Beta-ketosynthase domain (KS2): part of extender
FT
```

```
1824 aproletrArgProGlyThrGluInSerSerGlySerSer----- 1838
1839 -----GlySerGlyGlyGly----- 1844
1844 yGlyGlySerSerArgProAlaSerHisSerHis----- 1856
1857 -----AlaHisGlnHisSerProIleSe----- 1864
1864 rProArgThrGlnAspAlaLeuGlnInArg-----ProSerValLeuHisAsnThrG1 1882
1882 -----ACAGT 23444
23494 TCCGCCCGCACAGATCCAAATACCGCATCGTCGGCTTCCCTCCGG----- 1889
1889 ----- 1889
23383 TCCAAATGCGGCTCGGCGCTCGTCGATGTGTCCAGACGGGCGTCGACGTGCCGAGGTG 23324
1890 -----ValGluProSerLysProThrValLeuArgSerThrSerSe 1905
23323 AAGGCTGGGGGAAACCGTCCCAATCGACATCGGCCAGCAGGAGGTCTCATCGCGG 23264
1905 rProValArgPro-----AlaAlaThrPheProProAlaThrHisCysProIe 1921
23263 TCCAAATGCACTTGAAGTCCAGCAGCGCGCGCTCGCTTCATCGACGCCAGGCCATC 23204
1921 uGlyGlyThrLeuAspGlyVal-----TyrProThrLeu---Me 1933
23203 CGGTGACGTGTGCGCGGTTTCGCGGGCGCATCGCTCACCGGCCACCGTCCCCAC 23144
1933 tGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPro---Ar 1952
23143 GCACCGCGGTCCCGACCAATCCCTCGCGCGCGCTGCCGCAACACGCATCCAAGTAG 23084
1952 gAlaAspThrGlyHisAlaPhe-----LeuAlaLysProProAla----- 1965
23083 CGGTTCGACCGCGCATACGCACCTGTCCGCGCGCTCCCGCACACCGCCGACCGAGGAG 23024
1966 ArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuVa 1985
23023 AACATCAAAACGATCCAGTCCGAGT-----CGGCCAGCAAC 22985
1985 lProProValSer---GlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaPr 2004
22984 TCGTCCAGATTCCGCGCACCGACCGCCTTCGCGCGCGCTCCCGCACACCGCTCCAAAT---CC 22931
2004 oHisHisAlaSerPro-----AspProPro-----AlaProProAl 2016
22930 GCCAAACCCATCTCCATAGTCGCGCGCTTCGCGGACCGCGCGCGGTGTACCAACGCA 22871
2016 aserAlaSerAspProHisArgGluLysThrGlnSerLysPheSerIleGlnGluLe 2036
22870 GTACGGGACACTCCCGACCGATCCCTGCCAACACTCCAGCCACTGCGTCACGATCAGCC 22811
2036 uGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerGlyValGluProVa 2056
22810 ACGTCGACAGG-----GCACATCGACACCGCCAGCAGCCAGGCTCC 22769
2056 lSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluG1 2076
22768 AGCTCCGCCCGCAATCCC-----CAGCG 22745
2076 uLeuAspLysSerHisLeuGluGlyGluLeuArgProLys-----GlnProGlyProVally 2095
```

```
22744 CCGGAGGCTCGCTCTCGAGGCTGTCAACACCGCTGTCAAGCCCGCGCCCGCGGCC 22685
2095 sLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSe 2115
22684 AACACCGGGCAA-----CATGCGCCCGCAACACCGCGGTACACCCCGTCA---CCAA 22634
2115 rSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLe 2135
22633 ACCGTCCCCCGAGGAGCCACACCGCCGACCGGACCGGACCGACCGGCTTCCAAACGCG 22574
2135 uAlaGln-----HisIleSerGluValIleThrGlnAspTyrThrArgHisHisProG1 2153
22573 CGCACGAACACACACCGCGGAATGGCGACTTGAT-----CTCGCGCACCATCCG--- 22522
2153 nGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProVa 2173
22521 -----CCAAACACCGCCCAACCGGCGACCAATCGGTCTCCCGTCCGAT 22475
2173 lLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProProAspHisGlyAlaPr 2193
22474 GCGAGGTGATCAAGCCACCCAGAACTCGGGTGTCTCAGTCCCGCGACACGCGACGAC 22415
2193 oAlaArgGlySerProHisSerGluGlyLysArgSerProGluProAsnLysThrSe 2213
22414 CCGCAGACACCGGCGCT-----GCTCCGAGAC---ACGGTACATC 22377
2213 rValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrG1 2233
22376 T-----TCCACGCGACGCGCGCGCGGTCCACCGCCCA 22344
2233 uProGlyHisSerArgSerAla-----Valty 2242
22343 AAAAGGCGCTCGACGTCGCGCTTCGCGAAGCGCGCAGAGGTGCACAGTTCAGCGCGAA 22284
2242 rProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPr 2262
22283 ACCAGTT-----GGCAGCAGAGAGATCCCGGAGCGGCGGATTCATCGAGCCC 22236
2262 oGlyAsnThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMe 2282
22235 CAACAGCGMAAGCAGCCACCGAGTTGACCGGATCCGTTCAGCGCATTCGCGCGCGCTG 22176
2282 tValLysSerLysLys-----GlnGluIleAsnLysLeuAsnThrHisAsnArgAs 2300
22175 TCGACGCTCTCGCGGTCCGGTCTCTTCGACAGCGCCGATCCGACGACCTCGCGGCC 22116
2300 nGluProGluTyrAsnIle-----SerGlnProGlyTh 2311
22115 GCCTCCGACGATGACCTCGTCAAGCGGTGACCAATCCGTTCATCCGACGACCGCGCAC 22056
2311 rGluIlePheAsnMetProAlaIleThrGlyThrGly-----Le 2324
22055 GACCAACAGCAGGTACCGAAATCTTCGGGCGCGCGCACCTCCGAGAGGAAATGCCAACT 21996
2324 uMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThr----- 2338
21995 GATTCGGTACCGCGAGGAATCGACGAGCAAGCTTCGGAGTTCGGTTCGCCAGGAGGA 21936
2339 AsnMetGlyLeuGluAlaIle-----IleAr 2347
21935 CAACCGCGCAATACATCGCTCAACGATGCGCATGTGTCTGTCTCGCGCCCAAGATCCAG 21876
2347 glyAlaLeuMetGlyLysTyrAspGlnTyrGluSerProProLeuSerAlaAsnAl 2367
21875 ACGCGCGCAACAGTTGCCAGATCCCTTGTGTCAACCGCGGTCCAG---AACTGTTCGCG 21819
2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAla-----AlaMetProIleThrA 2385
21818 CACTTCGCGATGTCTGCGCGGTCTGCGCGCGCTGCGCGCGCTGCGACTCGATCCAGTAGTG 21759
2385 laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly-----GlyGlyLysA 2403
```


QY 772 hrLeuGlyAlaAspGlyProProGly-ProPro-----ThrProProArgArg 788
 Db 27700 CAAGTGTCTCGCGCGGTGTGTTCCCAACAACGCTTCCGGTGCCTCCACCTCGAACA 27641
 QY 789 ThrSerArgAlaProIleGluProThrProAlaSer---GluAlaThrGlyAlaProThr 807
 Db 27640 GGCATCACCAGTCCGCCAGACCTCACTGCTCGCGGAGATCGGTGCGGTCAACGCG 27581
 QY 808 ProProAlaPro-----ProSerProSerAlaProProProValPro 823
 Db 27580 AATCAACCGTGGCCACGAATCGCGCAGCGCGGTCCACTCGCACCAACGACGATT----- 27527
 QY 824 LysGluGluLysGluGluGluThrAlaAlaProProValGluGluGluGln 843
 Db 27526 CGTCATCGAAGTCAACGAGAGCGCGCCGCCACCGCGCTTTCAGTGGCGGAAGTT 27467
 QY 844 LysProProAlaAlaGluGluLeu---AlaValAspThrGlyLysAlaGluProVal 862
 Db 27466 CCACCCACGACGAGAACGCGCCCGGTTCGCTCTGA-----CCACCG 27419
 QY 863 LysSerGluCysThrGluGluAlaGluGlyProAla-LysGlyLysAspAlaGluAl 882
 Db 27418 ACGTCTCGGCTCGCGCGCATGCAGAACCGGCTCCAGAACGCGGGTGGACGCGAATC 27359
 QY 882 aAlaGlu-----AlaTh 886
 Db 27358 GATCAGACCCCGCTTCATCGCGCAATGCACTTCGGCGAACCTCCCGCGCTC 27299
 QY 886 rAlaGluGlyAlaLeuLysAlaGluLysGlyGlySerGlyArgAlaThrAl 906
 Db 27298 GTGCGCAGCACC-----GGCAGACATTAAGCCGCTCGTACGCGAAGCAG--TG 27247
 QY 906 aLysSerSerGlyAlaProGlnAspSerAspSerAlaThrCys-SerAlaAspGluV 926
 Db 27246 CCCTTCAAGGAATGGTAGACCCCTCTGCATCGATTTTCGTCGATTTCTCCGGGGCCA 27187
 QY 926 aLysGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuT 946
 Db 27186 TCGGTGAGCTCATGTTCTCGCGCGCCAGCGGTGGGAGAGCAGCCCGCTCGGTGTT 27127
 QY 946 hrProThr-----GlyAspProArgAlaAlaAsnAspSerProGlnLysProLeuAspLeuL 964
 Db 27126 CTTCCACTCGGTTCGATCGCTCGCAGGACGAGTAGAGCGCACCGCGCGCTCC 27067
 QY 964 ysGlnLysGlnArgAlaAlaAlaProIleProIle---GlnValThrLysValHisG 983
 Db 27066 GGATTTCCCGG-----GGGTCCCAACAGACCTGAACCTGACCGCTCCATG 27019
 QY 983 luProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProG 1003
 Db 27018 ATCCCGCAGGATCAGGGGCGCTTCCA-----AAGCCAGTTCTTCCACCGGCC 26971
 QY 1003 lnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysS 1023
 Db 26970 ACACCGAGCGCTCGCCAGCGTCCACACGAGTCCAGATCGCGGTGCGCGGAGCAGC 26911
 QY 1023 erArgSerProAlaPro-----ProAlaAspLysGluA 1034
 Db 26910 GATTTCCCGCAGCAGCCCGGTGATCGGACCAACGAGGATCGAACCCGCTGTCAACCGACC 26851
 QY 1034 laPheAlaAlaGluAlaGlnLysLeuProGlyAspProPro----- 1047
 Db 26850 GGTACAGCACCACTTCGCGCTCTGCCAAGCAACCAACCGCAGCAGCGGATGGTTC 26791
 QY 1047 ----- 1047
 Db 26790 CGCTCGGCAATCCGCGCGGTTCAGTACCCGACGCGCGGTGGATGCCAGCGATGA 26731
 QY 1048 -----CysTirThrSerGlyLeuProPheProValProProArgG 1061
 Db 26730 CCCTGCGGTGGAAGCGGTAGTGGCAGGTTCGCTTCGCTCCCGCGGTGCGGCGAA 26671
 QY 1061 luValileLysAlaSerProHis-----AlaProAspProSerAlaP 1075

Db 26670 AAAGGACCCAGTCCACCTCACCAACGCGGTGTGATCTCGCCAAAGCTGCCAGCA- 26612
 QY 1075 heSerTyrrAlaProProGlyHisPro---LeuProLeuGlyLeuHisAspThrAlaArg- 1093
 Db 26611 --CGGTTCGCGCTCGTTCGCGGTCTTGGCATACGCGGATCCCGGCCACCGCCGGA 26554
 QY 1094 ----ProValLeuProArgPro---ProThrIleSerAsnProProProLeuIleSerS 1111
 Db 26553 TTCGCGCACACTCTCTCGACCAACGTCGACACGCGCGTCCGACCAATTCGACAAAT 26494
 QY 1111 erAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerV 1131
 Db 26493 CGTGCAGCACCTCT-----GACCGACAAGCGCTCGA----- 26462
 QY 1131 alGlnLeuHisValProTyrrSerGluHisAlaLysAlaProValGlyProValThrMetG 1151
 Db 26461 -----CACCTCGCGGAACGCGGCTCAGCAGCTCAGCAGCCAGTACTCG-G 26411
 QY 1151 lyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluG 1171
 Db 26410 GAGTGCCTCATCTCGACCAACGCTCCAGC-----TCACCGATCAGGCTCGAGA 26363
 QY 1171 lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGly-ValProThrAlaGln 1190
 Db 26362 TG-----ATCGGACGTTCCGGTGGTGTAT 26339
 QY 1191 GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys 1210
 Db 26338 -----ATTCGCGGC-----CTCGGAGA----- 26321
 QY 1211 GlyIleProSerThrArgValProSerAspSerAlaIleThrTyrrArgGlySerIleThr 1230
 Db 26320 -----TCTGGCGAATCTCTCCAGCATCGCT---CCATATGGGCGGAATGGAAG 26273
 QY 1231 HisGlyThrProAlaAspValLeuTyrrLysGlyThrIleThrArgIleIleGlyGluAsp 1250
 Db 26272 CATCGACACCGCAACACC-----GGTCCGCGC 26243
 QY 1251 SerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrr 1270
 Db 26242 ACCCTGATGTGCAACCGATCAGCATCTCGGTGAGCAGCTCGCGGTCCACGAGAGCA 26183
 QY 1271 GluGlyLysLysGlyHisValLeuSerTyrrGluGlyGlyMetSerValThrGlnCysSer 1290
 Db 26182 CAACCGATTCCGGGCCAT----- 26165
 QY 1291 LysGluAspGlyArgSerSerSerGlyProPro----- 1301
 Db 26164 ---TCACGCGCGCATACCGATCCGTCGCGCCACCCATCCAGCAGAGGATCGACCTGAA 26108
 QY 1302 -----HisGluThrAlaAlaProLysArgThrTyrr 1311
 Db 26107 ACTCACCCTAGCAGCCGCGAGCATGGCACCCAGCAAGGCAACGCTTCATCAACC--- 26051
 QY 1312 AspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMet 1331
 Db 26051 ----- 26051
 QY 1332 GlyArgAlaIleProPro-----GluArgHisSerProHis-Hi 1344
 Db 26050 GGGCGGAGCCGCGCACCAACCGCGCGGCTCCGACACGACACACACACCGCGCATGCA 25991
 QY 1344 sLeuLysGlu-----GlnHisIleArgGlySerIleThrGlnG 1358
 Db 25990 CGCGAGCCAGCTCACCCACCGAGTCCGCCAACACACCATCCGG-----CCGAA 25943
 QY 1358 ylleProArgSerTyrrValGluAlaGlnGluAspTyrrLeuArgGlu----- 1374
 Db 25942 CACCCACGAAACCCAG-----CAGCGCAGCAAGCCCAATCTGCACCGCAACAAACCCGACT 25886
 QY 1374 ----- 1374

Db 29631 -----CTCAAGGTTTCCAGCGCACCTCCAGCAGCAACCGCTGCTGGCGATCCA 29582
Qy 174 ---GlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnIleSerLy 192
Db 29581 TCAGCAAGGCTACGCGGCTGATGCGCAACATGTCAGATCGAA----- 29536
Qy 192 sLeuLysLysGlnGlnGlnGlnGluGluGluGluAlaLysProGluProGluLy 212
Db 29535 -CTCCGCGC-----GTCCCGCAGAA 29516
Qy 212 sProVal-----SerProProProlleGluSerLysHisArgSerLeuVa 227
Db 29515 ACCCGCCTCACACATACGAGTCCGAGACGATCGG-----AT 29474
Qy 227 lGlnIleIleThrAspGluAsnArgLysLysAla-----GluAlaAlaHisArgI 244
Db 29473 CAGGATCGTAAACCGCGCAGATCCAGCGCGCATCGTCTGGGAAACCCAGCACCGCAT 29414
Qy 244 e---LeuGluGlyLeuGlyProGlnValGluLeuProLeuThrAsnGlnProSerAspTh 263
Db 29413 CCAGCGCGCGCCAGCGCGCAAGCTCTCCGCGCAGCACACCCCGGAAAC 29354
Qy 263 rArgGlnThrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuLeuTy 283
Db 29353 GACATCCCATCCGACAA-----CGCAACAAGATCTCTTCGG 29315
Qy 283 rPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAs 303
Db 29314 CACTGCTGCGCTGCTCACCGCGCTCGAGTGGAGCGGACCTCGCTCGCAG----- 29263
Qy 303 pGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArg 323
Db 29262 -TTCCTGATGAGAAACCGCGCCAGCGC-----CGCGGCGTGC 29225
Qy 323 gAlaLysGluSerLysValArg-----GluTyrTyrGluLysGlnPhePr 338
Db 29224 GATAGTGGAAACACCGAGTCCGCGCAACCGCACCCCGCGCAAGCGGTCC 29165
Qy 338 oGluIleArgLysGlnArgGluLeuGlnGluArgMet----- 350
Db 29164 GCAACTCCAGCGCAGAACAGATCAAAACCCAGATCAGGAAAGCAGCGTCCGCCAA 29105
Qy 351 -----GlnSerArgValGlyGlnArgLysGlySerGlyLeuSerMetSerAlaAlaArgSe 368
Db 29104 CCTGACCGCACCGAATAACCCAGCAGCGCGCAACGTG-----CT 29063
Qy 368 rGluHisGluValSerGlu-----IleIleAspGlyLeu----- 379
Db 29062 CCGCACCAACCAACAGTCCGCTCATCCGCGGCTGAGAAATGGCTGTTTTCCTCG 29003
Qy 380 -----SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaVal----- 394
Db 29002 CCACCGCGCGCAACAGGAACCAACCGGCTCAACAGCGCGGAATCGGTGCATCGTC 28943
Qy 395 -IleProProMetLeuThrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGl 414
Db 28942 CGTTCC-----GACCAGCAGCGCCT-----GTCCAAATCGAGCGG 28907
Qy 414 YLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGl 434
Db 28906 CAACGACGCAAGCCG-----GTCCTTCGCAACCGCGGATC-----GA 28868
Qy 434 uGlnGluLysGluThrPheArgGluLysPheMetGln---HisProLysAsnPheGlyLe 453
Db 28867 ACAGCGGAGTCCATCTCCGTGGAAAGTCTCGATGCGCGTCCGCGC----- 28819
Qy 453 uIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLeuTh 473
Db 28818 -----CAACCGCTCCACCGACCGCGCTC----- 28795
Qy 473 rLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGlyLysSe 493
Db 28794 -----CAAGCCACTCGTACACCGCTGGGTGGCGCCACAGTC 28757

Qy 493 rGlnGlnGlnGlnGlnGlnGln-----G 502
Db 28756 CCACAGCCACGACACCCAGCGAGCCACCGCGCGGCGATTAACGCGCCCAATGCATCCA 28697
Qy 502 nGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGl 522
Db 28696 GGAACGTTTCGCGCGCGCTAGTTCGCTTCGCGCACCCGCAATCCCGGAGAAG 28637
Qy 522 uLysGluLysGluLysGluAlaGlu----- 530
Db 28636 AAGAGAAAGAAAGAGAGACGACGATCCAGATCGCGGGTCAACTCTGCGAGATGCCATG 28577
Qy 531 -----LysGluGluGluLysProGluValGluAsn-AspLysGluAspLeuL 546
Db 28576 CCCCCGTACCTTCGGGGCAACACCGTGCCCATCCGCGCGGACAGACCCGATCA 28517
Qy 546 eLysGluLys-----ThrAspAspThrSerGlyGluAspAsnAspG 560
Db 28516 CTCGTCGCGCAGCACCCAGCGGTATGACGATCCCGCGAGCGCGCAGTCGACCGAA 28457
Qy 560 lLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGly----- 576
Db 28456 TGGCGGCGCAGCGTCTCCAGATCGGTGCGGTGAGCCACATCGCAAGCCACCACTCAA 28397
Qy 576 ----- 576
Db 28396 CCACGCGCGCGCGCGCGCTCATCCACCAACTCGTGGACTCCAGGCGCATTCACGC 28337
Qy 577 --ArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaI 596
Db 28336 CACGCGCGCTGCAACACCAACCGCGATCCCGTCTCTCCACCAAGGTGACGGGCAA 28277
Qy 596 leThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeu-----AsnGluSerA 614
Db 28276 CCACCTCCCCCAACGCGCGCGGACCGCGTCCCGTGTAGAGAGCGTCCCGTGTGAGATT 28217
Qy 614 rGtrPrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnT 634
Db 28216 GCGGGAGGAGCGCTCTCTCCGCGACCGTGTAGTGGCGC-----AACCGAGGCA 28169
Qy 634 rPserAlaIleAlaArgMet----- 640
Db 28168 CCAGCGCACCGCGCGCGGAGTGCCAGTCCCGTCTCTGCTCGCACGCGCGCGCGCA 28109
Qy 641 -----ValGlySerLysThrValSerGlnCysLysAsnPheTyrP 654
Db 28108 ACGCTGCCACGACTCAGGTGTGCGGTCCAGCTCCACGACGACGACGAGCCCGGGTTCT 28049
Qy 654 heAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLysMetG 674
Db 28048 CCGACTGGGTGAACTCAACACCCCGACGACGACCCACCGAGTCCCGGAGTCCG 27989
Qy 674 lLysGluArg-----AsnAlaArgArgLysLysLysAlaProAlaAlaAlaSerG 692
Db 27988 AATCCGAGTCGCGCGACACACCGCTGACGCTGACGACGACGAGTGGGATTTCGCGAAC 27929
Qy 692 luGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGlyValSerG 712
Db 27928 GCTCGTCCGACAGCCACGACTGCGAGAGCCAGACAGATCGGAGTCCGACTCTTGGGCGC 27869
Qy 712 lyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValP 732
Db 27868 GCGACGCGCAGGATCCACCTTCGACCTTCGCGCGGACATGCTCGCACCAACACCGAAC 27809
Qy 732 roArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleP 752
Db 27808 CCGGGGAGGAAAGCTCGAGTCCACGCGCGCACGGAACCTATCCAGGCTCGCGAAC----- 27753
Qy 752 roSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaT 772
Db 27752 -----CGTGGGTTCGCGGACCGGACGATTCGCGGCAACTGGGACCGCTCACCGAGATGG 27701

This invention describes a novel method nucleic acid (I) and its encoded polypeptide (II) containing at least one region that encodes an enzymatic activity in biosynthesis of spinosyns. (I) are used (i) to identify, inactivate or modulate genes involved in the biosynthesis of (II); (ii) to generate a library of polyketide synthases; (iii) for adding forosamine or trimethylrhamsone to a spinosyn or polyketide aglycone, and (iv) for recombinant production of the corresponding enzymes, which are used for production of (II), their precursors or derivatives, including production of transgenic plants that express (II) and thus have increased resistance to insects. (I) are also useful as markers for sequencing of the Saccharopolyspora spinosa genome. (II) are macrolides with insecticidal, but not antibacterial, activity, and can also be used to raise specific antibodies, useful for identifying expression clones in a gene bank. Cells transformed with (I) may produce (II) at significantly increased levels or produce new derivatives of (II). This sequence represents a genomic DNA fragment of the S. spinosa genome which contains the coding regions for proteins involved in forosamine and trimethylrhamsone biosynthesis

CC 38108 GCGCGGCGGAGATTGCTTCCAGTCGACCGGAAACAGGAGACTCGTGTTCATCGCACGG 38049
CC 2348 LysAlaLeuMetGlyLysTyrAspGlnTgpluGluSer----- 2360
CC 38048 GCGCGTCCCACTGCTTGTGCGAAATCTGGCGGAGCAGCAAGATCGATGGAGACGCG 37989
CC 2361 -----ProProLeuSerAla-----AsnAlaPheAsnProLeuAsnAlaSer 2374
CC 37988 GCTTGGCAGTCGTGCTGGTGTGAACACCGACACCGCATTCGGCGCCGCGGAAACGAGG 37929
CC 2375 AlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu 2394
CC 37928 TGCACCCGAGGAAGTGGCGGCTGCTTTGAGGACACACCTTCCACGCGAAGCGGG 37869
CC 2395 ThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
CC 37868 AGCCATCCCTCGTCGGATCTCT-----TGGGCTGGCCCGACGCGCTGCAGTGCC 37818
CC 2415 LysSer-----ProAlaProGlyLeuAlaSerGlyAsp 2425
CC 37817 GCATCCAGCAGCAGGCTGGATTCCGAACCTGCGCGCTTTCGCGGTGTGCGTGCGG 37758
CC 2426 ArgProProSer-----ValSerSerValHisSerGluGlyAspCysAsnArgArgThr 2443
CC 37757 AGGCGGACCTCGCGGAAGATCTCGCTCCGCGCTTCCAGCGCGCTGCAAC----- 37707
CC 2444 ProLeuThrAsnArgValTgpluAspArgProSerSerAlaGlySerThrProPhePro 2463
CC 37706 -----CCCTGGACCGCGCGCGGAAATCGAAGCCCGCTCAGCTGCGCGG 37662
CC 2464 TyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGly 2483
CC 37661 GCGTA-GAAGCTGTCCAGCGGAACCGCGACGACATCGTTGGCGCGGCCAAATCCCATCGG 37603
CC 2484 LeuProAlaGly-----SerGlyProLeuAlaGlyPro 2494
CC 37602 GCCCTGGCGCGGCTCTGCTGTGATCGCTGGAGTCAACGTGCGGTGCGGTGCGAGGTCCA 37543

RESULT 62
ID AAF88316/c
XX AAF88316 standard; DNA; 50000 BP.
XX AAF88316;
XX 28-AUG-2001 (first entry)
XX S. spinosa DNA fragment SEQ ID 5.
XX Forosamine; trimethylrhamsone; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; ds.
XX Saccharopolyspora spinosa.
XX DE19957268-Al.
XX 08-MAR-2001.
XX 29-NOV-1999; 99DE-01057268.
XX 27-AUG-1999; 99DE-01040596.
XX (FARB) BAYER AG.
XX Eberz G, Moehrl V, Froede R, Velten R, Salas JA;
XX WPI; 2001-267102/28.
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
XX recombinant production of insecticidal spinosyns and their derivatives.
XX Claim 7; Page 74-91; 354pp; German.

CC polypeptide (II) containing at least one region that encodes an enzymatic activity in biosynthesis of spinosyns. (I) are used (i) to identify, inactivate or modulate genes involved in the biosynthesis of (II); (ii) to generate a library of polyketide synthases; (iii) for adding forosamine or trimethylrhamsone to a spinosyn or polyketide aglycone, and (iv) for recombinant production of the corresponding enzymes, which are used for production of (II), their precursors or derivatives, including production of transgenic plants that express (II) and thus have increased resistance to insects. (I) are also useful as markers for sequencing of the Saccharopolyspora spinosa genome. (II) are macrolides with insecticidal, but not antibacterial, activity, and can also be used to raise specific antibodies, useful for identifying expression clones in a gene bank. Cells transformed with (I) may produce (II) at significantly increased levels or produce new derivatives of (II). This sequence represents a genomic DNA fragment of the S. spinosa genome which contains the coding regions for proteins involved in forosamine and trimethylrhamsone biosynthesis

XX Sequence 50000 BP; 6944 A; 14312 C; 19170 G; 9574 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.: 2,3e-10 Length: 50000
Score: 631.50 Matches: 687
Percent Similarity: 29.95% Conservative: 289
Best Local Similarity: 21.08% Mismatches: 1177
Query Match: 4.78% Indels: 1112
DB: 4 Gaps: 149

US-09-522-753-5 (1-2517) x AAF88316 (1-50000)

Qy 11 ThrTrpArgAlaThrGluProArgTyrProPro-His----- 22
Db 30142 ACGGCGGTCTGTTTCACGCGCTACCGCGCACCGCCAGACCTCGTGCCCATTCG 30083
Qy 23 ----SerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuG1 41
Db 30082 GCGGGCATCCGACACCGCTCCAG----- 30058
Qy 41 uTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnPr 61
Db 30057 ----CAACACGACCGCGCACCTCGCC-----CCACCGGTGCGCATC----- 30019
Qy 61 oGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnG1 81
Db 30018 -CGCGCGCGCGGACGACGACTTGACGCGCCCATCGGAGCAGACCCCGTTGACGGAGA 29960
Qy 81 uLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetG1 101
Db 29959 ACTCCAGAACGTCGTGCGGTGCGATCATCCGTGACGCCACCGCGGCGGAATTTCG 29900
Qy 101 uPhe----- 102
Db 29899 ACTCACCGCGCGGTGTTGACCGCGAGGTGTAACGCCACCAACGACGACGACACG 29840
Qy 103 -----IleGluSerLysArgProArgLeuGluLeuLeuPr 114
Db 29839 CCGTATCCACCGTCACCGCAGGACCCCTCGAAACAAACGATACGC----- 29794
Qy 114 oAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPro----- 129
Db 29793 -AACCGGACCGCAGCAGCGCTCCCGCATTCGCGTCCCGAGGTGCGCTTCGAAGCGCT 29735
Qy 130 -----AlaGlySerGluAsp-----LeuThrLysAspArgSerLeuTh 142
Db 29734 CCGGTGCTCTGTAATGAATCGGCGCCCATAGTCGTGTATCATCAAGCCCGGACACAC 29675
Qy 142 rGlyLysLeuGluProValSer-----ProProSerProProHisTh 156
Db 29674 CCGT-----CCGCTGCGGTGCAACGAGACGATCGATCCAGCCCG----- 29632
Qy 156 rAspProGluLeuGluLeuValProProArgLeuSerLysGluLeuLeu----- 173

```
Qy 1784 -----HisLeuThrLysProThrThrThrSerSerSerGluArgArgAsp 1799
Db 40006 CGAAGAAACCGGATCACTCGCGCATCTCGCAGAGAGCCACCCTCACACATACG 39947
Qy 1800 ArgAspArgGluArgAspArgAspArgGluArgGluysSerlleLeuThrSerThr 1819
Db 39946 AGTCCCAAGAC---GATCGGATTCGGATCGA-----ACA 39914
Qy 1820 ThrValGluHisAlaProIleTriArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db 39913 ACCCGCTGATCCAGCCAGCATCTCGCGGAATCTCGGCACCGCATCCAGCCCGCG 39854
Qy 1840 SerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln 1859
Db 39853 CCACCA-----AACGCCAAAGCTCTCGCGGAGGACACCCACCCCGGATAAC 39806
Qy 1860 HisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHis 1879
Db 39805 GACAGCCCA---TCCCCACGATCAACAACGGATCGTCAGCAGTCACCG----- 39761
Qy 1880 AsnThrGlyMetLysGlylleThrAlaValGluProSerLysProThrValLeuArg 1899
Db 39760 ---CCGCCACAGCCCGGTGCGGAGTCGGAGCTCCAGCCACCAATCAGTTGTGCA 39704
Qy 1900 SerThrSerThrSerProValArgProAlaAlaThrPhePro-----ProAlaThr 1917
Db 39703 GCAAGAACCGCAACCAACCGCGGTGTGCGATAGTCGAACACCGCAGTAGCGCGCAAC 39644
Qy 1918 HisCysProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeu 1937
Db 39643 GAACGCCCA----- 39635
Qy 1938 LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis 1957
Db 39634 -----GCACCCCGCAAGCCGGTTCGGAATCTCACACCGGACCAAGAACAA 39587
Qy 1958 AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLys 1977
Db 39586 -----AACCCAAATCCCGGAAG 39569
Qy 1978 GlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThr 1997
Db 39568 CACGATCGGCGCCGACCTCGACCCACCGC-----AATGCC 39533
Qy 1998 ProAlaLysAsnLeuAlaProHisHisAla---SerProAspPro----- 2011
Db 39532 CCAGCA-----CCGCCGAACATGCTCGCGCACCAACCCCAACAGCTGGGTTTCCT 39482
Qy 2012 ---ProAlaProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
Db 39481 GATCCAGTCCGGCAGCTCGCGACCGA---CCAACACAGACCGTTGACGTCGCGACTCG 39423
Qy 2031 PheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyThrHisGlySerSerTySerPro 2050
Db 39422 TTGACC-----GGCTCCGCGCGACGCGGCTGGCACACAGGTTCGCGCAACAC 39375
Qy 2051 GluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeu 2070
Db 39374 GCAGGCACACCGCTCGCGCGC---ATCAGCGCGCAGCGGCTTCAGTTCGAGTGGACCG 39316
Qy 2071 ---ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys 2089
Db 39315 AACCAGAAGCATTCGCGCGGAACCGCAACCGCATCTCGAA----- 39274
Qy 2090 GlnProGlyProValLysLeuGly----- 2097
Db 39273 CAGCCGAACCGCTCGCGCGGACATCGCGGATTCGGTGGCGCAATGCGGTACGGCC 39214
Qy 2098 -----GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer----- 2112
Db 39213 CGCCGCATCGAGCGCTCGGTCATCCCGTGGGTGCTGCCACCAAGCCCAAGCCAGCGA 39154
Qy 2113 -----GlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHis 2129
Db 39153 CACCGCAGGAGCCCGCCCGTCCGCGGTGCTGAGCCAGCCGCTCCAGAACCGGTTGCG 39094
Qy 2130 GlnArgValValThrLeu----- 2135
Db 39093 CGCGCGTAGTTTCCCTGACCCCGCAGCCCGCCGACCCCGGAGAACGAGGAGAAGACAC 39034
Qy 2135 ----- 2135
Db 39033 GAAGAACGAGAGATCCAGATCAAGAGTCAACTCGTCAGGTGCCACACCGCGGCGCTT 38974
Qy 2136 ---AlaGlnHisIleSerGluValIleThrGlnAspTyThrArgHisHisProGlnGln 2154
Db 38973 CGGCGCAACACCTTTGCCACGTGCTCTG---GTGTGTCAGCAGCTCGACACCCGCTCGCA 38915
Qy 2155 LeuSerAlaProLeuProAlaProLeu-----TySer 2165
Db 38914 GTGCCCGCGCG---CATGCACCACTCCGCGCAGCGGGAATCTGCCGAGATCGTCGTCA 38858
Qy 2166 PheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyIleu 2185
Db 38857 GCAGCGGTCCAACCTGGGCGCGATCCCGACATCGCAGCTCGCCACCTCGACCGAGCTC 38798
Qy 2186 Pro---ProProAspHis-----GlyAlaProAlaArgGlySer 2197
Db 38797 CCAGGCCACCAACTCATCCACGAGTCGGTGACCCCGGGGCGCTCCAGCGCGTCGCC 38738
Qy 2198 ProHisSerGlu---GlyGlyLysArgSerProGluProAsn----- 2210
Db 38737 CCGGAGCAGCAGCGCGGACACCGCGCTCCGCCACCAATGCCGGGCAACCAACCGCG 38678
Qy 2211 -----LysThrSer-----ValLeuGlyGly 2217
Db 38677 CCAGCAGCCCGTACCGCGGAATCAACACCGTCCATCCGACCGCGGACCTGGATCC 38618
Qy 2218 GlyGluAspGlyLeuGluProValSerProGluGlyMetThrGluProGlyHisSer 2237
Db 38617 GGCTGTCTCGCGGCAACACATCGCGCAGCGGAGGCA-----CCGAGCTTTCAC 38567
Qy 2238 ArgSerAlaValTyProLeuLeuTyArgAsp----- 2248
Db 38566 CCGCGCAACAGCCAGCTCGGTTCTCCGAACACACAGCGCGCGCGGAGCGACCGCATG 38507
Qy 2249 -----GlyGluGlnThrGluPro 2254
Db 38506 ATTGCGGGTTCCGTCACATCGACCAACACGAATCGTCCGGGTTCTCGGACTGCGCG 38447
Qy 2255 Ser-----ArgMetGlySer 2259
Db 38446 AACGCACAAACCCACGACGACGATGCACACAGGTTCGGAACACCCCTCGCGCGGACCTG 38387
Qy 2260 LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGlu----- 2277
Db 38386 TCAGCGCGCACACACGTCACAAACACACGAGGAGATCGCGGGAATCGCGCCCGCGCCA 38327
Qy 2278 -----SerAsnSerAlaMetValLysSerLys-----LysGlnGluIle 2290
Db 38326 ACCAATCTCGATCAGGTCAAGCATGCTGTGTGTCGCGGACGCTGCTCGGAAGCAACGT 38267
Qy 2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyArgAsnIleSerGlnProGly 2310
Db 38266 CCAAGTCTCGCGAGACGCGCACCGCATGCGCACCAACCA-----CGTCCCGGG 38219
Qy 2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGln 2330
Db 38218 GCG-----GGTCAGCCAAACCCGTCAGTTCGGGTGG---AGTTCGTCGCAAGTGCC 38169
Qy 2331 AlaValGlnGluHisAlaSerThrAsnMet-----GlyLeuGlu 2343
Db 38168 GAGCGCAGGCTCGGCGGAGTTCGTCTATCCGATCTGCTCCAGGAGACCGGCTTGCA 38109
Qy 2344 Ala-----IleIleArg 2347
```

| | | | | | | | | |
|----|-------|--|-------|----|-------|--|--------------------------------------|------|
| Db | 41964 | ATCGAAACCCGCAACCGCGCCGACCGAATCCCTTGCCCATCCAGCGCGCGCGATGTC | 41905 | Qy | 1487 | ValHisProLeuAspVal | -----MetAlaAspAlaAArgAlaLeuGluArgAla | 1502 |
| Qy | 1177 | InAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgG | 1197 | Db | 40933 | ACCAATCCACTTCCAGAGTAGTTTCATCCACATGCAAGTCGCGGAGTCCCGTGC | 40874 | |
| Db | 41904 | GTCAGACACTCCCGATCGCGGAGACACCGACTCAGGAG | 41857 | Qy | 1503 | -----CysTyrGluGlu | 1506 | |
| Qy | 1197 | lyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgV | 1217 | Db | 40873 | GCATGCCATCACCATTGTGATCACACGACGACACCGCGCGCTCGTATGACCAA | 40814 | |
| Db | 41856 | G-ACCGG-----GCGATCCCGCACCATCA---ACCCGATCGGCCAACACGGCGCA | 41807 | Qy | 1507 | ---SerLeuLysSerArgProGlyThrAlaSerSerGlyGlySerIleAlaArgGly | 1525 | |
| Qy | 1217 | alProSerAspSerAlaIleThrTyr-----ArgGlySerIleT | 1230 | Db | 40813 | TATTCAGATTACCGACCCCAACACGCGCGATCGTATCCCGCCCTCCCGT--- | 40757 | |
| Db | 41806 | GCTGCTCTCCAGAGCAGCCACCCGACATGTCACCGCAGGACGACCGTGCATCA | 41747 | Qy | 1526 | AlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAsp | 1545 | |
| Qy | 1230 | hrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluA | 1250 | Db | 40756 | -----AAGTGGCAATCAGAGCTCGCGCTCAATCGGAT | 40724 | |
| Db | 41746 | ACCGGCGACGACCGCC-----ACCATCCGAGCGCATCCGCGAACG | 41705 | Qy | 1546 | HisGlyAlaProPheAlaGlyHisLeu-----ProArgGlySerProValThr | 1561 | |
| Qy | 1250 | spSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleT | 1270 | Db | 40723 | CACCAAGCCTGGTTCAGTCCCATGCGCTCCACGCGCTCCACATCAGACCCGACAAAC | 40664 | |
| Db | 41704 | ACAAAC-TCGCGCGCAACCGCGCGGAACT-CGCGACCGAATGGCCGACCCAC | 41647 | Qy | 1562 | MetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGln | 1581 | |
| Qy | 1270 | yrGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCys- | 1289 | Db | 40663 | CAGCACTCGCCAAACGCT-----CGGTATCACCCGCTGCTCGAGGGCCCATTCG | 40613 | |
| Db | 41646 | ACCC-GGCGGACACCCACGACCCAGC-----AGGCTCAGCAAAACCACTGCA | 41597 | Qy | 1582 | AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro | 1601 | |
| Qy | 1290 | -----SerLysGluAspGly-----ArgSerSerSerGlyPro-----ProH | 1302 | Db | 40612 | -----GCGCGCTCAACCATTCGACG----- | 40592 | |
| Db | 41596 | ACGCGAACAACCGGACTGGGCGCCACAGCGTCCGATCCAAACGAGACCCGTCGGAACAA | 41537 | Qy | 1602 | GLUHIHISPro-----HisProLysSerProTyrGluHisLeuLeuArgGly | 1617 | |
| Qy | 1302 | isGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaI | 1322 | Db | 40591 | ---CACCATCTGATTACCGCGCTACCGCGTACCGCGACGCGCAGACCGCGTGCCTGC | 40535 | |
| Db | 41536 | ACACACATCCCGGACCCCGAGTCTGCTCCCGAGATGGCGTCCAGTCCGCGCATGCTT | 41477 | Qy | 1618 | ValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIle | 1637 | |
| Qy | 1322 | leSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGlu----- | 1338 | Db | 40534 | GC CGGG-----CATCCGACAGCGCTCCAAACACACC-AAGCGCGCGCTTCT | 40488 | |
| Db | 41476 | CGTCAACGCGTTCAG-GAACCACGGAAAAACCGAGTAGAGCGCCCGGCCCATCCCAAC | 41418 | Qy | 1638 | ProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro | 1657 | |
| Qy | 1339 | -----ArgHisSerProHisHisLeuLysGlu-----GlnH | 1349 | Db | 40487 | CCCAACCGGTGCA---TCCGAGCGCGCGCAACGACTTGACCGGCCCATCCGCGGCC | 40431 | |
| Db | 41417 | CACGTCCGCGCTGACCGCGCAACACGACCCACCGCCGACGAGACCGGGATCCCGAC | 41358 | Qy | 1658 | AspPro-----ThrTyrPro-----HisLeuTyrPro | 1666 | |
| Qy | 1349 | isHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluA | 1369 | Db | 40430 | AAGCCCCGTGACGCGAANAACCTCCAGCAACATCCCGCGCTGCCCATCCCGTCACACCA | 40371 | |
| Db | 41357 | ACCAC-----CCCAAGCTCCGCGACCGCGAGCCAGC | 41325 | Qy | 1667 | ProTyrLeuIleArgGlyTyr---ProAspThrAlaAlaLeuGluAsnArgGlnThrIle | 1685 | |
| Qy | 1369 | spTyrLeuArgArgGluAlaLysLeuLysArgGlyThrProProProProp | 1389 | Db | 40370 | CCGCAAGAGCCAGATCACACTCACAGACCGCAGTCTTGACCGCGCAGGTGCGAGGCC | 40311 | |
| Db | 41324 | GCTTCCAAACCCGACAAACGTTTCGCGGTCCGCTCCCA-GCACACCGCGCGCTCCTC | 41266 | Qy | 1686 | IleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGln | 1705 | |
| Qy | 1389 | roSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysp | 1409 | Db | 40310 | ACCAACACGCA-CGAAACA---CGCGGTATCCCGTCCCGCAGGACCTCGAACCACAA | 40255 | |
| Db | 41265 | CAACCGCGCGCGGAGGAGCAACGAAACG-----C | 41233 | Qy | 1706 | ArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAla | 1725 | |
| Qy | 1409 | ro-AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIle | 1428 | Db | 40254 | CGA-----ATACCGAACCCGACCCGAAACACACGCTCCCGCGTGC-CATTACCTA | 40205 | |
| Db | 41232 | CGCATCCGCGTGGAAATATCCGAGTCCACCCAAATAGGACAAATCGCTCGCTG | 41173 | Qy | 1726 | AlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValPro | 1745 | |
| Qy | 1429 | ProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlu | 1448 | Db | 40204 | GATAACCTTCGAAC-----CCT | 40187 | |
| Db | 41172 | AGCGCTGAGCGTCCGCGGTCTCCCGACAGATCAACGCGGACCGCTGGAGGTCCAC | 41113 | Qy | 1746 | ProThrPro----- | 1748 | |
| Qy | 1449 | GlySerIleThrGlnGlyThrProLeuLysTyrAsp----- | 1460 | Db | 40186 | CCGCGCGCTCGTGATGAAACCGCGCGGTAGTCTGCTGTCATCAGCCCCGCAACACGC | 40127 | |
| Db | 41112 | GGTCGATTCGATCCGAGCGGACCGCTCTGTCGCGAGTCTCGTGGTGTTCGAGGCTTG | 41053 | Qy | 1749 | -----GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln | 1766 | |
| Qy | 1461 | -----ThrGlyAlaSerThrThrGlySerLys---LysHisAspVal | 1473 | Db | 40126 | CGGTCCGCTCCCGCAACGAAAGCGATCAATCCCGCGCGCTCCACGCTTCCCAAG | 40067 | |
| Db | 41052 | TTCGAGGATGACGTGTGCGTGTGTGCTGCTGATCCCGACGACGACACCCACCGCGG | 40993 | Qy | 1767 | ProPheSerSerArgHisSerSerSerProLeuSerPro-----GlyGlyProThr | 1783 | |
| Qy | 1474 | ArgSerLeuIleGlySerProGlyArgThrPhe-----ProPro | 1486 | Db | 40066 | CGACTTCAGCAGCAACCGCTGCTGCGGATCCATCCGCAACGCTCAGCGGGGAATCC | 40007 | |
| Db | 40992 | AGGATGGCTGTCGCGCGCGCGGGGTGTCT-CCGTGAGGAGCTGGACATCCCGCGCG | 40934 | | | | | |

```
Qy 633 --AenTrpSerAlaIleAlaAargMetValGlySerLysThrValSerGlnCysLysAsnP 652
Db 43967 GCGACCAAGCAAGCGGGGCTCGCACCGCGCCTTCGCGAGGGCAACTGCGG---T 43911
Qy 652 heTyRPhenTyRlysLysArgGlnAsnLeuAapGluIle-----LeuG 667
Db 43910 TCCATCAGGCCCAACGACTCGGCAACGCGCACCGCATTCGGCGGTTCGTCGACGTCC 43851
Qy 667 lnGlnHisLysLeuMetGluLysGluAargAsn-AlaAargArgLysLysLysAla 686
Db 43950 GCCAGCACTACCTCCCGGGTTCACCTCGCGCGCACGCAACACCCCAACCGC 43791
Qy 687 ProAlaAlaSerGluGluAlaAlaPheProProValValGluAapGluMetGlu 706
Db 43790 GCATGTGCCAGCTCTGATCGACCGTCACTCGGAGCCGTCGTC-----ACC 43746
Qy 707 AlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAla 726
Db 43745 GCTGCACACGTCAACACACGAGAGTGCCTCGGCCAACCGCGGCTCAGGCCAACCGCT 43686
Qy 727 -----SerGlyAsnGluValProAargGly 734
Db 43685 TGGAGCAGACCAACGCTCTGCTGGGTCTCCGCAAGTGGCGTGGCGGTGCCGGATCC 43626
Qy 735 GluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPro 754
Db 43625 GCCTGTCTCT---CGCGGCGAGTAGACCAACGACATCAGGAACAGCGCGCTGGAATCAAGT 43569
Qy 755 HisThrGluAlaAla-LysAspThr----- 762
Db 43568 TTCTCGGCTAGCGCGGCAGATCCGCGAGGGTCCGCGGGAATCCGTTGCGAGTTTGATG 43509
Qy 763 -----GlyGlnAsnGlyProLysProAlaThrLeuG1 774
Db 43508 AGATCGTCTTTACCGCAGGAGTAGTATCACCAGCAGGGCCCGACCGGTTCGTCGAC 43449
Qy 774 yAlaAspGlyProProGlyProProThrProProAargThrSerArg-----Al 792
Db 43448 TGCCTCGCCACCGGCTCCAGGTCACTCGAAACAGCGCGTCCGCGACAGGCTCCCGATT 43389
Qy 792 aProIleGluProThrPro-----AlaSerGluAlaThrGlyAl 805
Db 43388 GCCAGCTGTCCGACGACACGCTTCGTAGCTGAGAGATCGATCGAGGGCAGCAACCGC 43329
Qy 805 aProThrProProAlaProPro-----SerProSerAlaPr 818
Db 43328 CCACTCTCATCCATCAGCGCACCGATATCGCTCGGGTCCCTGGTCTCCAGCGGCACC 43269
Qy 818 oProProValProLysGluGluLysGluGluThrAlaAlaAlaProProValG1 838
Db 43268 CTG-CTGATCGTCCCTCGCGAGCGGAA-----AGTTCCAGCGCGGTGAA 43225
Qy 838 uGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAl 858
Db 43224 TGAGAACGGCAGCCAGCGCGCCCTTGCTCC-----GG 43192
Qy 858 aGluGluProValLysSerGluCysThr---GluGluAlaGluGluProAlaLysG1 877
Db 43191 GCGGTCCGCGAGCGGAGCGGCTCGAGCGCGCATCCAGCAACGCTGGTTCGACAGGTA 43132
Qy 877 yLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysG1 897
Db 43131 TCCCGACGCC---ATCGCATCGGCTCGCGCGGGCGGATTTTCGCGCGAACCCTCGTC 43075
Qy 897 uGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAsp----- 914
Db 43074 ACCAGCGCCATCGCGCTTGAACCCCTGGAACAACGCGCCGCTACTCGAACCCGTTTCC 43015
Qy 915 -----SerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
Db 43014 GCCCAGACGCGAGTAGAAGTATCATCGAGGAGATCGACTCGGCACCCCTCAGCCGACCATC 42955
Qy 929 a-Glu-----GlyGlyAspL 934
Db 42954 AGGAACCGCCCTTACCACCGGTGGTGGCGACGACGAGTTCCCGGTGGCGGCGTCCA 42895
Qy 934 yAsnArgLeuSerProAargPro-----SerLeuLeuThrProThrGlyAsp--- 950
Db 42894 ACCCCACCTCTCGACCCACCGCGCGGGGGGAATACAGACGACACCCCTTCGCCGACCGCC 42835
Qy 951 -----ProArgAlaAsnAlas 956
Db 42834 TTCGTGGGGTGGCTTACCACGACGAGTGGATTTTCGATGGCGTTCGTTCGCCACACGCTC 42775
Qy 956 erProGlnLysProLeuAspLeuLysGlnLysGln----- 968
Db 42774 TTCGCGAGTTTCAATTGAATCTGTGTCTCCCGCAGGGGCCACGACTACCGGACCCGCG 42715
Qy 968 ----- 968
Db 42714 CGCGCGTTTCGGGACACACGCGGTGTTTCCAGGTGAGCTCTTCCACCCGCTCACAGCC 42655
Qy 969 -----ArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluP 984
Db 42654 GAGACGCGCCCTGCGTGCAGGGCCATCTCCAGATGGCGTGGCGGGACGACGACCTC 42595
Qy 984 roProArgGluAapAlaAlaPro---ThrLysProAlaProPro-----A 998
Db 42594 GCCCAGCAGCGATGCTGGGCCACGACGATGCTCCCGACCAACCGCGCGGTTCAG 42535
Qy 998 laProPro-----ProProGlnAsnLeuGlnProGluSerAspAlaProGlnG 1014
Db 42534 CACATCCCATCACCATCGGCCAGCGACCGCCCGCCAGACGAGTGTCTCCGCGCG 42475
Qy 1014 lnProGlySerSerProAargGlyLysSerArgSerProAlaProAlaAspLysGluA 1034
Db 42474 AGCCAGCGCATACCGTGCATACATCGC---CGTAATCGATGTTTCCAGCCAGTACCCTG 42418
Qy 1034 laPheAlaAlaGluAla-GlnLysLeuProGlyAspProProCysThrThrSerGlyLeu 1053
Db 42417 CCCTCGAAGGCATACGTCGCGCACTC-----GACCGCGCTGC----- 42379
Qy 1054 ProPhePro-ValProProArgGluValIleLysAlaSerProHisAla-ProAspPro 1073
Db 42378 -----CCCGTCCCG-----GCGAATACCGCGCACCATTCGACCGGAAACACC 42337
Qy 1073 erAlaPheSerTyRAlaProProGlyHisPro-----LeuProLeuGlyLeuHisA 1090
Db 42336 ACGGTGTGGCGCGCGCCACCGCGGCGACACAGTTTCGGTTCGTTCCGTACGGTCTTACG 42277
Qy 1090 spThrAlaArgProValLeuProArgProProThrIleSerAsnProPro----- 1106
Db 42276 CAGCATCGGAATCGGCATGCGCTGATCAGCCCAAGATGCAATCGGGAATCAGGGCCGA 42217
Qy 1107 -----ProLeuIleSerSerAla--- 1112
Db 42216 CAGGTCCGATCAGGGCCGAGTTCGAGATCGTGGGACCCCTTGTCTGGCGGACCCCG 42157
Qy 1112 ----- 1112
Db 42156 GACGCGTCCGGAAGCGGACGCGCTCTCGAACCTGACGACCAACCAATACTCCGCGTAGC 42097
Qy 1113 -----LysHisProSerValLeuGluArgGlnIleGlyAlaIleS 1126
Db 42096 CATCACACCTGCGGTGTCGAGCTCACCCGTCACCGTCAACGTCGACACCGCGGTAGGT 42044
Qy 1126 erGlnGlyMetSerValGlnLeuHisValProTyRSerGluHisAlaLysAlaPro--- 1144
Db 42043 -----CCCGACCGCGGTAGTCCAGCT 42022
Qy 1145 -----ValGlyProValThrMetGlyLeuProLeuProMetAspProLysL 1160
Db 42021 GCGTGCAGTTTCGCGGAACCTCTGCAACATCGGCT---CCATCGGATACGAATGAAACGC 41965
Qy 1160 ysLeuAlaProPheSerGlyValLysGlnGlnLeuSerPro-----ArgGlyG 1177
```

US-09-522-753-5 (1-2517) x AAF88313 (1-50000)

QY 16 GluProArgTyrProPro-----HisSerLeuSerTyrProValGlnIleAla 31
 Db 45889 GAACCA-----CCGCAACACCGATGCCATTCGCGGAGCATCCGACACCGTCTCA 45836

QY 32 ArgThrHisThrAspValGlyLeuLeuGluTyrGlnHis---HisSerArgAspTyrAla 50
 Db 45835 ACAACACCAAC-----CAGCACCTCTACCCCAACCGGTATCCG- 45798

QY 51 SerHisLeuSerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSer 70
 Db 45797 TCCGCGCTCTCCGGAACAGCTTGCACGCGCGTCCGCTGCCAGGCCCGC-CTGACCGGA 45739

QY 71 GluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSer 90
 Db 45738 GAATTC-----CACAAACGTCCCGGCGTCCGACAT 45709

QY 91 TyrLeuProGluLeuGlyLysSerGluMetGluPhe----- 102
 Db 45708 CACCGTCACACCGGCAAGGCGAGATCGCATTCGCGGAGCAACGACGATGACCCGC 45649

QY 103 -----Ile 103
 Db 45648 CAATGCAACGCCCAACAGGAGCAACAGCGAGTATCCACCGTCACCGCAGGACCTC 45589

QY 104 GluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu 123
 Db 45588 GAAACCAACAGTAGTCG-----AATCGCGCGGAGCGACATCCCGAGC 45544

QY 124 LeuAla-----ThrGlyGlnProAlaGly 131
 Db 45543 ACTGCCATTCGCGAGATACCCCTCGAAACCTTCGCGGCGTCCGCTGGCGAACCGCGGCC 45484

QY 132 SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProPro 151
 Db 45483 ATAGTCGTGGTACATCAGCCCGCGGAAACACACCGGTCCG-----ACTCCCGCT 45436

QY 152 SerProProHisThrAspPro---GluLeuGluLeuValProProArgLeuSerLysGlu 170
 Db 45435 CAGCGAAACCGGATCGATCCCTCGCCGCTTCATGCGCTCCCGAGGACACTCCAGCAACAG 45376

QY 171 GluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnIle 190
 Db 45375 CCGTGTGCTG-----CGGATCCATCCG-----CAAGCGCTC 45346

QY 191 SerLysLeuLysLysGlnGlnLeuGluGluGluAlaAlaLysProProGluPro 210
 Db 45345 ACGCGGGAATCCCGAAGAAACCGCATCGAACCGATCGCGGTCCGTGAGGAGGCGCC 45286

QY 211 GluLysProValSer---ProProProIleGluSerLysHisArgSerLeuValGlnIle 229
 Db 45285 TTCCCGCAGGTAGTCTTCCACGCGCTTC----- 45256

QY 230 IleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGly 249
 Db 45255 -----CGGTCCGATCGAACCAAGCTCGCATCCAGCGCGCATCGC 45211

QY 250 -----ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265
 Db 45210 TGGGAAGTCGCGACCGCATCCACGCGCGCGGCAACCAACGCGAGCTCTCCGCGGA 45151

QY 266 TyrHisGluAsnIleLysIle-----AsnGlnAlaMetArgLysLysLeuIleLeuTyr 283
 Db 45150 GGACACCTGACCGGGAACCGCATCCCATCCCGATATCGCAGCAGATCGTCTTCGCG 45091

QY 284 PheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAsp 303
 Db 45090 GACCGGTGCGCTCGTCTGCTACCGCGCGTGGCGT-----CGTAGCGAT 45046

QY 304 GlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgTg 323
 Db 45045 CTGCTGTCAGTTCCTG-----GTGCAAGAACCGGCGCGCCCTCGGCGTCCG 44995

QY 324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhePro---GluIleArgLys 342
 Db 44994 ATAGTCGAACACCGAGTCGCGGCAACCGCACCCCGAGCACCCTCGGAGCGGTTCG 44935

QY 343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgLysSerGlyLeuSer 362
 Db 44934 CAACTCCACACAGACAGCAATCAAAACCCAGATCCCGGAAAGCGCGTCCGACCCGAC 44875

QY 363 MetSerAlaAlaArgSerGluHisGluValSerGluIleAlaAspGlyLeuSerGluGln 382
 Db 44874 GTCAGCGCGCTCGAATACCCAGCAGCGCGCGCATCTCCCGCACCCAGCGAGCAG 44815

QY 383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
 Db 44814 GGTGCGCTCGTCGCGAGCTCGGTCCGTAGC-----CCGCTATTACGACGATGG 44761

QY 403 AspGlnGln-ArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422
 Db 44760 GGGTACCATGTCGCTC-----AACATGCGGCGAGCGAGCGCGCGC 44716

QY 422 lTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgG 442
 Db 44715 CTGCTCGCGCAGGAAAGTCTGTCCAACCTCGCC-----GGAACCTACCGCTGC 44668

QY 442 uLysPheMetGlnHisPro-----LysAsnPheG 452
 Db 44667 CCGTTCAGCGCGCAGCGCGCATCGAACAGCGCAATCCCTCTGTTGGTGCACACGCGG 44608

QY 452 YLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLe 472
 Db 44607 GTTGGCGCGCGCATCTGTCTCGGTGACCTC-GCTGAG-----CGTCT 44564

QY 472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGlyLys 492
 Db 44563 CCGTATCCCGCTCGGTCTCCACAGCGCGCGCGAGCGACACCGAGCGGAGTCCCT 44504

QY 492 sSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 512
 Db 44503 GTGACCGCGGTAATGCGCAATGCGTCCAGGAACGCTTCGCGCGCGCATTAATGCGCT 44444

QY 512 tProArgSerSerGlnGluLysAspGluLysGluLysGluLysGluLysGluLysG 532
 Db 44443 GACCAGGAGCGCGCGCAACACCGGAAACGAGGAGAACACACCGAGAACACAGGTCG 44384

QY 532 uGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLys-ThrAspA 552
 Db 44383 TATCAGCATGCTGTCGATTCACGCGCGCGCGCTTCGCGCGCGAACACCGCTC 44324

QY 552 spThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer----- 566
 Db 44323 CCACGTCTCGGTGACACAGCACTCGATCACCGCTCGGCGCGCGCGCGCGCATGCA 44264

QY 567 -----LysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyA 581
 Db 44263 CCACCGCGCAACACGGAATTCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 44204

QY 581 rgIleThrArgSerMetAlaAsnGluAlaAsnSer-----G 593
 Db 44203 GATCCGCGACGTCGACCGCGCACCGCATCGACACCGCGCGCGCGCGCGCGCGCGCG 44144

QY 593 luGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer- 612
 Db 44143 CCAACTCGGTGATTCAGAGGACATCCAGCGCACCGCGCGCGCGCGCGCGCGCGCG 44084

QY 613 -----SerArgTrp-----ThrGluGluG 619
 Db 44083 CACCGGTTCGTCACAGGTGGGGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44024

QY 619 luMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg----- 632
 Db 44023 AAATCAGCACGTTGCGTTCGCTC-GGG---ATCCGCGCAACACCGCGCGCGCGCGCG 43968

Db 37669 CGTTGCCGATCAGCGGTGCGCCGGTGAGGGCTTGGGTGGGCTCGTTGATCGCGTTGAGGA 37728
Qy 2139 le-----SerGluValIleThrGlnAspTyrThrArgHisHisProG 2153
Db 37729 TTTGTTGCTCAGGGTGTGAGTGGCTGTGGCGGGGGCGTTGAATCCGCTCTCGACTA 37788
Qy 2153 lnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProV 2173
Db 37789 GTAGCTCGCGCTTAAGCG 37845
Qy 2173 allLeuAspLeuArgArgPro---ProSerAspLeuTyrLeuProProAspHisGlyA 2192
Db 37846 CACTACCACCGTTTACCGCATTTGCGGATCAGACACCGCGCTTCCCGCGCGCGCGCGTTCG 37905
Qy 2192 laProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysT 2212
Db 37906 CGCGTTGCGCGCTTGCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37952
Qy 2212 hrSerValLeuGlyGlyGlyGluAspGlyLeuProValSerProGluGlyMetT 2232
Db 37953 -----CCGTTCCCG 37977
Qy 2232 hrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluInt 2252
Db 37978 ACCCG 38037
Qy 2252 hrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProAlaPheP 2272
Db 38038 CG 38087
Qy 2272 heSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnLysLeuAsnL 2292
Db 38087 ----- 38087
Qy 2292 ysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrG 2312
Db 38088 -----CCGCGGAGCG 38121
Qy 2312 luilePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlav 2332
Db 38122 CGATCCACCGCACCGCG---CCACCGTCCGGGCTG----- 38156
Qy 2332 alGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleAlaArgLysAlaLeuMetG 2352
Db 38156 ----- 38156
Qy 2352 lysTyrAspGlnTrpGluSerProProLeuSerAlaAsnAlaPheAsnProLeuA 2372
Db 38157 -----GATCGCGCGCTGCC-GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38192
Qy 2372 snAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspH 2392
Db 38193 AGGTGAGCATGCGCGGTGCG 38252
Qy 2392 isThrLeuThrSerProGlyGlyGlyGly-----LysAlaLysValSerG 2407
Db 38253 GTCCCGGCACATACCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38312
Qy 2407 lyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgP 2427
Db 38313 CCG 38372
Qy 2427 roProSerValSerSer--ValHisSerGluGlyAspCysAsnArgArgThrProLeuTh 2446
Db 38373 AGCGCGCGTTTCGCGCGGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38427
Qy 2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPr 2466
Db 38428 ---CGCGC-CGCGCACCGACCGCGCGGTGCGGAGACACCGCGCGGTTCGCGCGCGCGCG 38482
Qy 2466 oLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAl 2486
Db 38483 CCG 38542

Qy 2486 aglySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLe 2506
Db 38543 GGCCCGAACAGCATGCT 38602
Qy 2506 u 2506
Db 38603 G 38603
RESULT 61
AAF88313/c
ID AAF88313 standard; DNA; 50000 BP.
XX
AC AAF88313;
DT 28-AUG-2001 (first entry)
XX S. spinosa DNA fragment SEQ ID 2.
XX Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX Saccharopolyspora spinosa.
OS DE19957268-Al.
PN 08-MAR-2001.
PD 29-NOV-1999; 99DE-01057268.
PF 27-AUG-1999; 99DE-01040596.
PR (FARB) BAYER AG.
PA Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
PI WPI; 2001-267102/28.
DR New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives.
XX Claim 7; Page 31-49; 354pp; German.

CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (ii); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhannose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC genome which contains the coding regions for proteins involved in
CC forosamine, trimethylrhannose and polyketide synthase biosynthesis

XX SQ Sequence 50000 BP; 6867 A; 14165 C; 19274 G; 9694 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.3e-10 Length: 50000
Score: 631.50 Matches: 680
Percent Similarity: 32.10% Conservative: 317
Best Local Similarity: 21.89% Mismatches: 1171
Query Match: 4.78% Indels: 958
DB: 4 Gaps: 139

| | | | |
|----|-------|---|-------|
| Qy | 1127 | nglyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyPr | 1147 |
| Db | 35106 | GGGG-----GTGCGG----- | 35126 |
| Qy | 1147 | ovalThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa | 1167 |
| Db | 35127 | GATCCCGCGGGGAAGCCCTGGCCAACTCCGCGGTTCGCCGCGCGCCGCC--GGAGCCGA | 35185 |
| Qy | 1167 | LysGlnGlnLeuSerProArgGlyGlnAla-----GlyProProGluSe | 1183 |
| Db | 35186 | AGACCGTGGCGGTGGCCCC--GGGCGCGCTTGCCTGCCGCGTGGGAGAGCCGATC | 35242 |
| Qy | 1183 | rLeuGlyValProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeuGlySerVa | 1203 |
| Db | 35243 | CGCGCGCGCGCGGAGCGCGCGAGCCGAAGAGCAGCCAGCGTTCGCCGCGCGCGCGC | 35302 |
| Qy | 1203 | lProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIl | 1223 |
| Db | 35303 | C-----GGCGCGCGCTATGCGCGCGCGCGTGAAGTAGTACCGCGTCCACCGATTCCGCCGG | 35359 |
| Qy | 1223 | eThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIl | 1243 |
| Db | 35360 | CGCGCGCGCGCGGCC----- | 35375 |
| Qy | 1243 | eThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLe | 1263 |
| Db | 35376 | ----- | 35395 |
| Qy | 1263 | uProLysGlyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGl | 1283 |
| Db | 35396 | TGCGCGCGCGCGCGCGTCCCGCGCGCGCAGC-----GCTGTGTCGCCGCG | 35443 |
| Qy | 1283 | yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGl | 1303 |
| Db | 35444 | TGCGCGCTTCCCGCTGGCCGA-----CA | 35470 |
| Qy | 1303 | uThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSe | 1323 |
| Db | 35471 | GCGCGCGCGCGCGCGCGCGCGCGCTCCGCGAAGCTGCTGCGCGAACCGCGCAT | 35530 |
| Qy | 1323 | rSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHi | 1343 |
| Db | 35531 | GCGCGCC-----GGTGC CGCGCGCGCGAAGCAGCGCGC-CAGCGCCACCGCGCCCA | 35580 |
| Qy | 1343 | sHisLeuLysGluGln-----HisIleIleArgGlySerIleThrGlnGlyIlePr | 1360 |
| Db | 35581 | CGGCGCGCGCGGAGCTGCGGCGCGCGCGCGTCCGCGA-----CC | 35622 |
| Qy | 1360 | oArgSerTyrValGlnAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysAr | 1380 |
| Db | 35623 | CGCGCGTGGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTGACATG | 35682 |
| Qy | 1380 | sGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGl | 1400 |
| Db | 35683 | CCGTTC-GTGAATCCGCGCGCGCGCGCGCGCT----- | 35712 |
| Qy | 1400 | nAlaLeuGlyProLeuLysLeuAspAlaHisGluGlyLeuValAlaThrValLysGl | 1420 |
| Db | 35713 | -----CCGCGG----- | 35718 |
| Qy | 1420 | uAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPr | 1440 |
| Db | 35719 | -GCGCGAAGAGCAGCGCGCGCGGTGCG-----CCAGCCCGCGC | 35756 |
| Qy | 1440 | oLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAs | 1460 |
| Db | 35757 | GGCGCGCGCGCGC-CCGCGCGCGTGAAGGCTACTACGCGCGCGCGC----- | 35798 |
| Qy | 1460 | pThrGlyAlaSerThrThrClySerLysLysHisAspValArgSerLeuIleGlySerPr | 1480 |
| Db | 35799 | -GGCGCGCGCGCGCGCGCGCGCGCGAAGCATGCGCTTCCGCGC-----GGCTCCGC | 35851 |

Qy 1676 rAlaAlaLeuGluAsnArgGlnThrIleLeuAenAspTyr-----IleThrSerGlnG 1694
 Db 39731 AGCGCGCTGTTTCAGCGCCCAACCGCGTTCGCGCGTTCACCGCGTGA 39790
 Qy 1694 nMetHiShiAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSe 1714
 Db 39791 GCGCGAGCGCGCTGTCGCGCGTTCGCGCGCGCGCGCTCCCGCTCCCGCGC---GG 39847
 Qy 1714 rProArgGluSerSerLeuAlaLeuAsnTyrAlaAla-GlyProArgGlyIleAlaAspL 1734
 Db 39848 TGCGCGCTTTCGCGCGCTTCGCGCATTCGCGCGTTCACCGCGTTCGCGCGTGA 39901
 Qy 1734 eUerGlnValProHisLeuProVal-----LeuValProProThrProG 1749
 Db 39902 TGCGCGCTTTCGCGCGCTTCAGGCGCGTTCAGGCGCGTTCGCGCGCTTCGCGCGTTCGCGCG 39961
 Qy 1749 lyThrProAla-----ThrAlaMetAspArgLeuAlaTyrLeuProThrAlaP 1765
 Db 39962 CGACCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 40021
 Qy 1765 roGlnProPheSerArgHisSerSerProLeuSerProGlyProThr-His 1784
 Db 40022 CGCGCGCGCGCATGCGCGCTGTTAGCGCGTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCG 40081
 Qy 1785 Leu----- 1785
 Db 40082 CTTTTCGCGCATCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGT 40141
 Qy 1786 ThrLysProThrThrThrSerSerGluArgGluArgAspArgAspArgGluArgAsp 1805
 Db 40142 ACCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCG 40201
 Qy 1806 Arg----- 1806
 Db 40202 CGTTACCAACCGTACCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGT 40261
 Qy 1807 -----AspArgGluArgGlySer-SerIleuThrSerThrThrThrValGlu-- 1822
 Db 40262 CGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGT 40321
 Qy 1823 -----HisAlaProIleThrArgProGlyThrGluInse 1834
 Db 40322 CGTTGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCG 40378
 Qy 1834 rSerGlySerSerGlySerGlyGly-----GlyGlySerSerSerArgProAl 1852
 Db 40379 CAGCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCG 40438
 Qy 1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAsp----- 1869
 Db 40439 CTTGCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCG 40498
 Qy 1870 -----AlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleI 1887
 Db 40499 CGCTGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCG 40558
 Qy 1887 eThrAlaVal-GluProSerLysProThrValLeuArgSerThrThrSerSerProV 1907
 Db 40559 TGCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCG 40612
 Qy 1907 aArgPro-----AlaAlaThrPheProProAlaThrHisCysProLeuGlyG 1923
 Db 40613 TCGCACCTTCATGCTGGAATGGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCG 40672
 Qy 1923 lyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu----- 1937
 Db 40673 CCACATTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 40732
 Qy 1938 -----LeuProLysGluAlaProArgValAlaArgp 1948
 Db 40733 CGTCACCACTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 40786
 Qy 1948 roGluArgProArgAlaAspThrGlyHisAlaPhe-----LeuAlaLysProProAlaArgS 1967

Db 40787 CATTGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 40846
 Qy 1967 exGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProp 1987
 Db 40847 CGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 40906
 Qy 1987 roValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisA 2007
 Db 40907 CG-----TTGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCG 40948
 Qy 2007 laSerProAspProAlaProProAla-----SerAlaSerAspProHisArgG 2024
 Db 40949 CGACGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 41008
 Qy 2024 luLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyH 2044
 Db 41009 CGTTACCAACCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 41046
 Qy 2044 isGlySerSerTyrSerProGluGlyValGluProValSerProValSer----- 2060
 Db 41047 -----GGGTGCGCGCGCGCACCTTCGTTGCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 41083
 Qy 2061 -----SerProSerLeuThrHisAspLysGly-----LeuP 2071
 Db 41084 CGCTGCTGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 41143
 Qy 2071 roLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnP 2091
 Db 41144 CACCGCGCGCGCGCACCTTCG 41203
 Qy 2091 roGlyPro-----ValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgP 2108
 Db 41204 CGCACCTTCGTTGACG 41263
 Qy 2108 roLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysG 2128
 Db 41264 CTTGCGCGCGTCA-----CGCGCTTCG 41320
 Qy 2128 lyHisGlnArgValValThrLeuAlaHisIleSerGluValIleThrGlnAspTyrT 2148
 Db 41321 GCGAACCG 41340
 Qy 2148 hrArgHisHisProGlnGlnLeuSerAlaProLeuPro-----AlaProLeuTyrSerP 2166
 Db 41341 -----TCGAATCGCTTCG 41395
 Qy 2166 heProGlyAlaSerCys-----ProValLeuAspLeuArgArgProProS 2181
 Db 41396 CGCGCGCGCTTCGCGCGCTTCG 41452
 Qy 2181 erAspLeuTyrLeuProProAspHisGlyAlaProAlaArgGlySerProHisSerG 2201
 Db 41453 CGCGCGCATTCGTTGCGCGCTTCG 41505
 Qy 2201 luGlyGlyLysArg-----SerProGluP 2209
 Db 41506 CGCGTGAACCGCGCATGACCG 41565
 Qy 2209 roAsnLysThrSerVal-----LeuGlyGlyGlyG 2219
 Db 41566 CGCGCATGACCGCGCATTCG 41625
 Qy 2219 luAspGly----- 2223
 Db 41626 CCGATGCGCGCATCGCGCGCTTCG 41685
 Qy 2223 luProValSerProProGluGlyMetThrGluProGlyHisSerArg-----S 2239
 Db 41686 GCACCGCGCGACG 41745
 Qy 2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSer----- 2255

Qy 2001 -----AsnLeuAlaProHis-----HisAlaSerProAspProProAlaProAla 2016
Db 2868 CGCTCCACACAGCGACCCACATCCCGAAACCCCGAACCTCCCGACATCCCTC 2809
Qy 2016 aserAlaSerAspProHisArgGluThrGlnSerLysProPheSerIleGlnGluLe 2036
Db 2808 CTCACGTCAGCGCAAAATCGCTCC----- 2784
Qy 2036 uGluLeuArgSerLeuGlyTyrHisGlySerSerTyr---SerProGluGlyValGluPr 2055
Db 2783 -----CGTTCMAACGATAGTCGCGAGCTCCACCCCGATCCATCAGCGACTCGCG 2731
Qy 2055 oValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuG1 2075
Db 2730 AGCAGTCCCTCGAACACACCGGCC-----CACTCAACCGCGTCCAGC-CAGGAACA 2678
Qy 2075 uGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro----- 2091
Db 2677 ACTCCGCGAGCGCGGTATGACCGACCGTGCCTCCGCTCGTCCGCGCGAGCGCGGGA 2618
Qy 2092 ----- 2092
Db 2617 TGGCGCGCGCGGTGCACTGAGCGAGTCCTGTGCGAGGGCGGACGAGCGTCCGTCGGGGC 2558
Qy 2092 yProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSe 2112
Db 2557 CGATTTCGAGCGAGGTGTGAGCGCCCTGTTCTGAAGCCATGAGATCCGTCGCGCAAC 2498
Qy 2112 rGlnProSerSer-----SerProLeuLeuGlnThrAlaProGlyValLysGlyHi 2129
Db 2497 GGACCGTGCTCGCGGCGTGTTCGACCCAGTAGTCCGGGTGCACATGCTCGCGCGGGA 2438
Qy 2129 s----- 2129
Db 2437 GGGCGCGCGCGTGACGTTGAGAGACGAGGGAATCCCGGGGCGCTGAAGTGACCTGCT 2378
Qy 2130 -GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrAr 2149
Db 2377 CGCGCGCGCGCGGAAGTCGCCCAACATGGCGTCCATGTG----- 2338
Qy 2149 gHisIleProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAl 2169
Db 2337 -CGCGAGTGGAGCGGTGGTGTGTCGCGACCG-----CCGGGTGC 2297
Qy 2169 aserCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeu----- 2185
Db 2296 GCGCGCTCGTGCCTCCCATTTGCTGCGGAGGTCCAGGACCGCGTCTCTGTCCTCCCGGAGA 2237
Qy 2186 -----ProProAspHis-GlyAlaProAlaArgGlySerProHisSerGluG 2202
Db 2236 GGACGATCGACCGCGCGCGTTACCGCGGCGTGGCGGACCCCGGATGCGTATTCGTCG 2177
Qy 2202 yGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyI 2222
Db 2176 GCAGCGGAGG-----ATCTCGTCTCCGACGCCCTCA 2144
Qy 2222 leGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValT 2242
Db 2143 TGGCCACCATGGCTCCGCGCGAGCGG-----ACCCCTTG-CATCAGCGCGCTCGTCG 2091
Qy 2242 yrProLeuLeuTyrArgAspGlyGlnThrGluProSerArgMetGlySerLysSerP 2262
Db 2090 ACCACGATGCCCGCGTCCGCAAGGACAGCATCCGGGACATCGCGGCGCGC----- 2035
Qy 2262 roGlyAsnThrSerGln-ProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
Db 2034 -----CAGTTCACCGAGGATGCCGAGCGGTAGTCGGCGCTCAGACCC 1989
Qy 2282 MetValLysSerLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
Db 1988 CAGGTCTCCAGCGCGGACAGCGCCACCTCGAAGGCGGACAGGGCGGGCTGGCGGAAA 1929

Qy 2302 ProGluTyrAsnIleSerGlnProGlyThrGlu-----IlePheAsn 2315
Db 1928 CCGCTGTCTCGATCAGCGCGCTTCGGGAGAGTCTCTGCGGTGCGAAGAGTACGTCCCGC 1869
Qy 2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHis 2335
Db 1868 AGCCACAGGGACCGGGGTGCGTGGCGGCGGTCTCCGCGAGATCTCTGTCGATG 1809
Qy 2336 AlaSerThrAsnMetGlyLeuGluAlaIleAArgLysAlaLeuMetGlyLysTyrAsp 2355
Db 1808 GCCTGGCGGAAGACGGGTACGCTCGTACAGTTCGGGCGCCATCGCTCGCGCG----- 1755
Qy 2356 GlnTrpGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
Db 1754 ---TGGGTTCCTCCGCGCGGGAAGAT---ACGGCGAGTTTCGCC-----GAGGTG 1710
Qy 2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
Db 1709 GTTCGTCTCTGACGACGCGCGGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 1650
Qy 2396 SerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArg-LysAlaLys 2415
Db 1649 AGGAACCTCGTCGCGGTCTCGGCCACGACCCACGACGATGCTCGAACACCGCGCTCC 1590
Qy 2415 sSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSe 2435
Db 1589 GACACCAAGCGCGCGCGACCCCGCGGACTCACCCTCCACCATCCGACCCCGCAC-- 1532
Qy 2435 rGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSe 2455
Db 1531 -----CAACCGCCCAACCCCGCGCAACC-----GACGCGCTGC 1497
Qy 2455 rSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVa 2475
Db 1496 CCGCGCAACCACTCCGACCGCGCGCACCA-----CC 1461
Qy 2475 lMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHi 2495
Db 1460 CACGCGACCAACCCCGCAACCCGACACCCCGCGGACCCCAACTCTCTGACGCGCGCGCA 1401
Qy 2495 s---HisAlaTrpAspGluProLysPro 2504
Db 1400 CCGCCACCGCGCGCGCGCGCGCTCTCTCCA 1370

RESULT 59
AAI99682_37
Continuation (38 of 45) of AAI99682 from base 3700001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000

| | | |
|----|------|--|
| Qy | 791 | -----ArgAlaProLeGluProThrProAlaSerGluAlaThrG 804 |
| Db | 6729 | GAGGGAGGTGCACCGCGCAGCTAGGCCGCTCGCATCTCGCCCTGCGAGTCCCGGCCAC 6670 |
| Qy | 804 | lyAlaPro-----ThrProProProAlaProProSerProSerA 817 |
| Db | 6669 | CGCGTCGCGGTGTACACCGTAGAGCGCCAGAGCGCGCCAGAGACCATCACCGCAA 6610 |
| Qy | 817 | laProProValProLysGluGlySgLuGluGluThrAlaAlaalaProProv 837 |
| Db | 6609 | GAGACGGGTGGAGCATCGACCGCTTGCAAGGG- GGTGCGTCGGTGGCGCGCCA 6551 |
| Qy | 837 | aLGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAla----- 852 |
| Db | 6550 | GGAGCTCGAGGAGACACCGTCCAGGTACGGTTCCAGGGCTTGGCGCAGTCGACATCT 6491 |
| Qy | 853 | ----ValAspThrGlyLyAlaGluGluProValLysSerGluCyThrGluGluAlaG 871 |
| Db | 6490 | GCTGGCGGAAGACCGGT-----GAGAGCGG---AGGAGTTCCTGCGCCATGCTTCCC 6440 |
| Qy | 871 | lu-----GluGlyProAlaL 876 |
| Db | 6439 | AGTGGGTGCCCTGTCTCCGAAACAGCATGGGATCTTTCCGTGGCGCGCGTCCGGCCA 6380 |
| Qy | 876 | ysGlyLyAspAlaGluAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysL 896 |
| Db | 6379 | CACCTGTACACCCCCCGGTGGTGCCTCGGCCAGTGGCTGAGTGGCTGCAGGA 6320 |
| Qy | 896 | ysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerA 916 |
| Db | 6319 | ACTGTCGCGGTCTCTCGGCCACGACACCGCAGTCTCGAACACCGACCGC---TCCG 6263 |
| Qy | 916 | spSerSerAla----- 919 |
| Db | 6262 | ACACAAAGCCCGCGACCCACCGGAGCTACCCCCACACATCCGACCGCCACCAA 6203 |
| Qy | 920 | -----ThrCysSerAlaAspGluVala 927 |
| Db | 6202 | CCGCCACAACCCACGCAACGACGCGCTGCCCCCGCAACGCGCACTCCGACGCGCGC 6143 |
| Qy | 927 | spGluAlaGluGly-GlyAspLysArgLeuLeuSerProArgProSerLeuLeuThr 946 |
| Db | 6142 | ACACACCCACGGCACCAACCCCCGAAACCGACACACCCCGCGACCCA-----ACT 6092 |
| Qy | 947 | ProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro----- 960 |
| Db | 6091 | CCTGACGCGCGCGCACCCCACTCCGCGCCCCCGAGCGGCTCTCCAAATCATATGCG 6032 |
| Qy | 961 | -----LeuAspLeuLysGlnLysArgAlaAlaAlaLeuProProIle 976 |
| Db | 6031 | CATTGTCCTCACTACCCCGNACGACACACCCCGCAGCGCGAGCGGACCTCCACCC 5972 |
| Qy | 977 | GlnValThrLysValHisGluProProArgGluAspAlaLaProThrLysPro----- 994 |
| Db | 5971 | CCGGCCACTCCACCTCATTCGCCAACACACGAAACCGACCACTCGACCAATCCACCTGCG 5912 |
| Qy | 995 | -----AlaProProAla----- 998 |
| Db | 5911 | ACGAGGCTCATCCACATGCAACGTCCGCGGCAACACCCCGCGCGCAACGCCATCACCA 5852 |
| Qy | 999 | -----ProProProGln-----AsnLeuGlnPro 1007 |
| Db | 5851 | TCTTGATCACCCCGCCACACCCGACGAGCGTGGCATGCCCGATGTTGCACTTCAACG 5792 |
| Qy | 1008 | GluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArg----SerPro 1026 |
| Db | 5791 | ACCCCAACCAACCGCGCGTCTCCCGGAAACGACCTGCCATAAGTGGCGAGCAACGCT 5732 |
| Qy | 1027 | AlaProProAlaAsp----LysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAsp 1045 |
| Db | 5731 | GCGCTCGATCGGATCAACCAACGCTGACCGTCCGCGTGCC----- 5690 |

| | | | | | |
|----|------|---|------|-----|------|
| Qy | 1046 | ProProCysTrpThrSerGlyLeuProPheProValProProArgGluValLysAla | 1065 | ACA | 5660 |
| Db | 5689 | CTTCCA-----CCACATCCATCCGCACGG----- | | | |
| Qy | 1066 | SerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuPro | 1085 | | |
| Db | 5559 | ACCCGCACAGCCAAACGCTGCGGATCACCGCTGCTCGGACGGACCATTCGGCGCCG | 5600 | | |
| Qy | 1086 | LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro | 1105 | | |
| Db | 5599 | --TCAACCCATTCGACGACCGCTGCTGATTACCGGCACTCCCGCGCA-----CCA | 5552 | | |
| Qy | 1106 | ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIle | 1125 | | |
| Db | 5551 | CCGCAAAA----- | 5543 | | |
| Qy | 1126 | SerGlnGlyMetSerValGln-LeuHisValProTyrSerGluHisAlaLysAlaProVa | 1145 | | |
| Db | 5542 | CCGATGACACGACGATTACGCTCGGACAGCCGCTCCACCAACAGACACCCACACCT | 5483 | | |
| Qy | 1145 | lGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheS | 1165 | | |
| Db | 5482 | CGGCCACCGCAC-----CCCATCGGCCCGACCCGCTGACGCTTCACCG----- | 5437 | | |
| Qy | 1165 | rglyValLysGlnGluLeuSerProArg----- | 1175 | | |
| Db | 5436 | -GCCGTCCGCGCAGACAGACCCCGCTGCCGCGAGAACTCCACAAACGACCCGGGCTCGACA | 5378 | | |
| Qy | 1176 | -----GlyGlnAlaGlyProProGluSerLeuGlyValProTh | 1188 | | |
| Db | 5377 | TCACCGTCAACCCCGCCAAACGCGCAGCAACACTCCCGCGACCTCAACGCTTGACACG | 5318 | | |
| Qy | 1188 | rAlaGlnGluAlaSerValLeuArgGlyThr-AlaLeuGlySerValProGlyGlySerI | 1208 | | |
| Db | 5317 | CCGATCGAGCGCCACCAACGACGACGAAACGCGGTATCCCGTCAACCGCGGACCT | 5258 | | |
| Qy | 1208 | leThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlyS | 1228 | | |
| Db | 5257 | CGAGCCCAAGGTGTAGCGACGCGTCC-GGATGTGACGCTGCT-----GGAC | 5211 | | |
| Qy | 1228 | erIleThrHisGlyThrProAlaAspValLeu-----TyrLysGlyT | 1242 | | |
| Db | 5210 | AGGCCGTCATGCGGTAGCCCTCGAGGTCTCTCGGTGGCCGCGCAGAGGTGGCGGTAG | 5151 | | |
| Qy | 1242 | hrIleThrArgIleIleGlyGluAspSer-----ProSerArgLeuAspArgGlyA | 1259 | | |
| Db | 5150 | TCTGTACTGCATCGCGCGGAGACACCGGTGCTGGACCCGCGCA-----CGTGCGC | 5097 | | |
| Qy | 1259 | rgGluAspSerLeuPro-----LysGlyHisValIleTyrGluGlyLysLys- | 1274 | | |
| Db | 5096 | GGGTGCGATCCCGCCGCTCCACGCGCTCCAGGACACCTCCAGCATCAACCGTGTGTCG | 5037 | | |
| Qy | 1275 | GlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGly | 1294 | | |
| Db | 5036 | GGATCCATCCCAACGCGCTCACGCGGACTG-----ATCCCGAATAAACCCCGCA | 4989 | | |
| Qy | 1295 | ArgSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMet | 1314 | | |
| Db | 4988 | TCGAACTCCCGCGCACCTCCAGGAAA-----CCGCCCCGCGCGTATACGAC----- | 4941 | | |
| Qy | 1315 | GluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAla | 1334 | | |
| Db | 4941 | ----- | 4941 | | |
| Qy | 1335 | IleProProGluArg-HisSerProHisLeuLysGluGlnHisIleArgGlySe | 1354 | | |
| Db | 4940 | ---GAACCCCGCCCGCGCTCCGGATCATAGAAAGCCCTCCACGTCCCAACCCCGGT-- | 4886 | | |
| Qy | 1354 | rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgG | 1374 | | |
| Db | 4886 | ----- | 4886 | | |
| Qy | 1374 | uAlaLysLeuLeuLysArgGluGlyThrProProProProSerArgAspLeuTh | 1394 | | |

1684 hrilleleAsnAspTyrile---ThrSerGlnGlnMetHisAsnThrAlaThAlaM 1703
 3929 CCCTCCACACCCCGTATCCACAACTCCGCGCATGACACACACAC--- 3880
 1703 etAlaGlnArgAlaAspMetLeuArg-GlyLeuSerProArgGluSerSerLeuAlaLeu 1722
 3879 -----CAGCGGGCA-----CTCGCGGGAACGCGCGACCGCAACTCCGCGCAACGCC 3831
 1723 AsnTyrAlaAlaGlyProArgGlyleileAspLeuSerGlnValProHisLeuProVal 1742
 3830 TCACGGTCCGCGCATACACGCCACACACAC---ACCGCGCGCCCAAG---CCCATC 3777
 1743 LeuValProProThrProGlyThrPro---AlaThrAlaMetAspArgLeu---AlaTyr 1760
 3776 AAGTCCGCTCGAGTTCCTGACATCCCTGGCGCTCTCCCGCGCTCGGCTCACACGAGC 3717
 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSerProGly 1780
 3716 AGGTGTCGCGCCACCGCGGCCCATCCACCGGGCGACGTGCGCACCACTCCCGGTG 3657
 1781 GlyProThrHisLeuThrLysPro-----ThrThrThrSerSerSer 1794
 3656 CTCGCGTACGAGTACGTGCGCGGGCGCCACTCCCGCTCCGCGACGCGCTCTCTCC 3597
 1795 GluArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSerIle 1814
 3596 AACGGCGCCGACCAACCGCGCACAAACGCGCCGGAAGACCGCACGCGCAAACTCACTC 3537
 1815 LeuThrSerThrThrThrValGluHisAlaProIleTyrArgProGlyThrGluLysSer 1834
 3536 TCACCCCTCCCGCCACACCGCCAGCACACCTACC---AACCCATCCACACACCGCTCA 3480
 1835 SerGlySerSerGlySerSer-----GlyGlyGlyGlyGlySerSer 1848
 3479 TCACGAGTTCGCGACATCAACACCGCCACCGCGGTGCGTCCGCGCCCGCCACC 3420
 1849 SerArgProAlaSerHisSer-----HisAlaHisGlnHisSerProIleSerProArg 1866
 3419 ACACGCGCCCGCCACACACACACCGCGCGCGCCCGCCACACAGCATCCGCGCCCGCC 3360
 1867 ThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIle 1886
 3359 ACCGACGCGCC-----CGCGCGTTCACACACACACGCGCGCGCCCGCC----- 3318
 1887 IleThrAlaValGluProSer----- 1893
 3317 ---ACGCGCATATCACCGCGCTGCACCAACCCACACAGCGCCACTCCCGCTGCAGC 3261
 1894 -----LysProThrValLeuArgSerThrSerThrSerSerProValArgPro 1909
 3260 ACGCCACTCCCGCGCCACACAGGAGAGACACCGCGCGCGCCCGCCACATCGACCGCC 3201
 1910 AlaAlaThrPhePro-----ProAlaThrHisCysProLeuGlyGlyThrLeuAs 1926
 3200 TCACGAGGTGGCGGCCAACACTTCCCTGCTCACACACCGCGCTT----- 3155
 1926 pGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAl 1946
 3154 -----CCACCTCCACCGCAACCA---CTCGCGCGCCACACCGCTCCACCCCTCC 3108
 1946 argProGluArgPro----- 1951
 3107 GCCACCATCAACCGCGCGCGCTCGCCCTCGGACACACACCGCGCGCCGACAGC 3048
 1952 -----ArgAlaAspThrGlyHisAlaPheLeuAlaLysProPro 1964
 3047 CCCCTACACCGCGCCGAAACCGGTGCGCACACCTCCCGGATAGCGCCACCGCTCCACC 2988
 1964 oAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLe 1984
 2987 ACTTCACGCTGCTGCTACCGCGCGCCCACTTCCTCCCAACCGCGCACACCGCGCCAGC 2928

1984 uValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys----- 2000
 2927 GAGGCCCCCTCA-TCCACCCCAAGGAGCGATCCACACCCCGCATCCACACACTCGAC 2869
 2001 -----AsnLeuAlaProHis---HisAlaSerProAspProProAlaProAla 2016
 2868 CGCCTCCCAACACGCGACACCCACATCCGCGAAACCCCGGAACCTCCCGAGATCCCTC 2809
 2016 aSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLe 2036
 2808 CTCACGCTCCAGCAAAATCGCTCC----- 2784
 2036 uGluLeuArgSerLeuGlyTyrHisGlySerSerTyr---SerProGluGlyValGluPr 2055
 2783 -----CGCTCAACGCATACGTGCGAGCTCCACCCCGCATCCATCCACCGCTCGCG 2731
 2055 oValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuG 2075
 2730 AGCAGTCCCTCGAACACACCGCGC-----CACTCAACCGCGCTCCAGC-CACGAACA 2678
 2075 uGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro----- 2091
 2677 ACTCCCGCAGGCGGTATGACCGACCGTCCCTCCGCTCGCGCGCGCGGGA 2618
 2092 ----- 2092
 2617 TGGCGCGGCGGTGACCTGAGCGAGTCTGCGAGGCGCGACAGGTGCGCGTCCGCGGC 2558
 2092 yProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSe 2112
 2557 CGATTTTCAGGCGAGTGTGACCGCTGTCCTGAAGCCATGAGATGCGTCCGCGAACC 2498
 2112 rGlnProSerSer-----SerProLeuLeuGlnThrAlaProGlyValLysGlyHi 2129
 2497 GGACCGTGTGCGGCGGTGTCAGCCACAGTAGTCCGGGTGCACATGCTCTCGCGGGA 2438
 2129 s----- 2129
 2437 GGGCGCGCGGTGACCTGAGACGACGCGGAATCCCGGGCGCGTGAAGTGACCTGCT 2378
 2130 -GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrAr 2149
 2377 CGGCGCGCGCGGGAAGTCCGCCCAACATGGCGTCCATGTG----- 2338
 2149 gHisIleProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAl 2169
 2337 -CGGCGAGTGAAGCGGTGCTGCTGCGCAGCG-----CCGGGTGC 2297
 2169 aSerCysProValLeuAspLeuArgProProSerAspLeuTyrLeu----- 2185
 2296 GCGCGCTCGTCCCGCCCATTTGCTGCGCGAGGTCCAGGACCGCGTCTCTGCTCCCGGGA 2237
 2186 -----ProProAspHis-GlyAlaProAlaArgGlySerProHisSerGluG 2202
 2236 GCACGATCGACCGCGCGCTTCCCGCGCGGTGCGGACCCCGGATGCGTATTCGTGCG 2177
 2202 lyGlyLysArgSerProGluProLeuSerValLeuGlyGlyGlyGluAspGlyI 2222
 2176 GCACGCGGAGG-----ATCTGCTCTCGGACGCGCTCGA 2144
 2222 leGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValT 2242
 2143 TGGCCACCATGCTCCCGCGGACGG-----AGCCCTTG-CATCAGCGCGCTCGTGGC 2091
 2242 yzProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerP 2262
 2090 ACCACGAGTCCCGCGTCCGCAAGGACGAGCATCCCGGCGACATGGCGCGCGCGC----- 2035
 2262 roGlyAsnThrSerGln-ProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
 2034 -----CAGTTACCGCGGAATGCGCGAGACGAGTAGTCGGCGCTCAGACCC 1989
 2282 MetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301

[illegible]

| | | |
|------|--|------|
| 4886 | ----- | 4886 |
| Qy | uAlaLysLeuLeuLysArgGluGlyThrProProProProProSerArgAspLeuTh | 1394 |
| Db | ::: ::: | |
| 4885 | -----CGACGGAAACTCTCCACCGCATCCCGACCGGACCGACGATTC | 4839 |
| Qy | rGluAlaTyLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLe | 1414 |
| Db | ::: ::: | |
| 4838 | CACAAGT-----CCTCGCGCACTCCACACCCCGGAAACGGCAC | 4797 |
| Qy | uValAlaThrValLysGluAlaGlyArgSer---IleHisGluLeProArgGluGlu-- | 1432 |
| Db | ::: ::::: | |
| 4796 | GCCATCCCCCAATTCGAATTCGGCTCGTCAACGTGCAGCGACGCGCGGACCGGAGGA | 4737 |
| Qy | -----LeuArgHisThrProGluLeuProLeuAlaProArgProLe | 1446 |
| Db | | |
| 4736 | GCAATGTCACGCTGGCCCGCGCCCTCTCAGCTGTTCTTGAGGTATCCGCCGACG | 4677 |
| Qy | uLysGluGly-----SerIleThrGl | 1453 |
| Db | | |
| 4676 | -GCGAGGAGTGGGGTAGTCGAAGATCAGCGTGGTGGCGAGGAGGCCGGTGACGGC | 4618 |
| Qy | nGlyThrProLeuLysTyRAspThrGlyAlaSerThrThrGlySerLysLysHisAspVa | 1473 |
| Db | ::: ::: | |
| 4617 | GTTGAGCGGTTGCGCATTCGACGGCGCTCAGCGAGACGAAGCCAGGTCGCGGAAGC | 4558 |
| Qy | LatGserLeuIleGlySerProGlyArgThrPheProProVal--HisProLeuAspVa | 1492 |
| Db | | |
| 4557 | TCGCTCA-----GGCGTAGCGCGTGGGGTCTGTGTCGAGCACCGT | 4513 |
| Qy | lMetAlaAspAlaArgAlaLeuGlu-----ArgAlaCysTyRgluGluSerLeuLy | 1509 |
| Db | | |
| 4512 | CGCGCGTACGTACGGACCAAGTCGAGAGCGCAGCGTCTCTGTCGCGCGTGCCATGGC | 4453 |
| Qy | s---SerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaProVa | 1528 |
| Db | ::: ::: | |
| 4452 | CTTGAGCCGT-----CGGAGAACGATCGGGGATGCGGTGGCGGTGTCGAGTCCGGT | 4399 |
| Qy | lIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyRgluAspHisGlyAl | 1548 |
| Db | ::: | |
| 4398 | GGTT-----TCCGGGGAGCGTGCTTCGGG-GATGTCGCTGATGAGGGGCG | 4352 |
| Qy | aProPhe---AlaGlyHisLeuProArgGlySer-----ProValThrMetArgGluPr | 1565 |
| Db | | |
| 4351 | AGAGTCGGGAGCGGGAGGAGTTGCGGTGAATCGTCCAGTCGA-TGTCGGCGACC | 4293 |
| Qy | oThrProArgLeuGlnGluGly-----SerLeuSerSerSe | 1577 |
| Db | ::: ::: | |
| 4292 | GTACACAGGTCGTCTCATGTTCCACGCTGGCCAGTGCACACGCGCGTCTCCGCGC | 4233 |
| Qy | rLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluLeAlaLysSerProHi | 1597 |
| Db | ::: | |
| 4232 | GTACATCCCGCCAGACCCGACGCGCGATCTGCCCCACGCGCCCTCCGCCATCCCCCA | 4173 |
| Qy | sser---ThrValProGluHisHisProHisProIleSerProTyRgluHisLeuLeuAr | 1616 |
| Db | | |
| 4172 | CCAGCCACCGACCCCAACCGCACCCCGGACGGCCCTCACAC----- | 4124 |
| Qy | gGlyValSerGlyValAspLeuTyRArgSerHisIleProLeuAlaPheAspProThrSe | 1636 |
| Db | ::: ::: | |
| 4123 | -----GCCGTGCCGAACGATTGCTCCACAT-----ACGCGTTC | 4089 |
| Qy | rIleProArgGlyLeProLeuAspAlaAlaAlaTyRLeuProArgHisLeuAl | 1656 |
| Db | | |
| 4088 | GCGCGCGGTAATTCCTCTGTCCG-----CGGCCCCGAACGTC | 4050 |
| Qy | aProAsnProThrTyRPro-His-----LeuTyRProProTyRLeuIleArgG | 1672 |
| Db | | |
| 4049 | GCGCAGCCGACGAGAACACACCAACGACCCCGAAGATCCGCCCCCGTCATCATGC | 3990 |
| Qy | lyTyRProAspThrAlaAlaLeuGlu-----AsnArgGlnT | 1684 |
| Db | | |
| 3989 | AGATTCCAGCGCGCAGCGCTTCGCGCGCAGCACCCCGTGACACGCTCGGACGACAAC | 3930 |

| | | | |
|----|------|---|------|
| QY | 459 | -----GluArgIysThrValAlaGluCysValLeuTyr-Tyr-TyrLeuTh | 473 |
| DB | 7635 | GGATGGGTAAAGACACAGGTCCAGCGGTCTCGGCTGTGCG----- | 7593 |
| QY | 473 | rLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGlyLysSe | 493 |
| DB | 7592 | -----GCGGGGTCCGAG | 7582 |
| QY | 493 | rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln----- | 507 |
| DB | 7581 | TCCGGCGAGCAGTCCGCGAGCATGCCGCGCACCGGTGTCCACGCTCCGACAACTG | 7522 |
| QY | 508 | -----GlnGlnGlnProMetProArgSerSerGlnGlnGluLysAspGluIy | 523 |
| DB | 7521 | ACCACCCACTGCCAGAGACCCCGGAAACCCGCGGGCGCCCTCCACATCCAGCCAAA | 7462 |
| QY | 523 | sGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGl | 543 |
| DB | 7461 | GCGETCTCGTCCAAACGATACCTCGCGACGTCCACCCGCGACACCGACCGCCGG | 7402 |
| QY | 543 | uAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGl | 563 |
| DB | 7401 | ACCAACACTCCCGAACACACCGGACCACTCAACCGCCACCCACCCACCGAACACTCGGC | 7342 |
| QY | 563 | aValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleTh | 583 |
| DB | 7341 | CACGGACATCAGGAAGGTCCGAAGCGCTTCGCC--CGCGCGACAGATCCGACGAC | 7285 |
| QY | 583 | rArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnSerAlaGl | 603 |
| DB | 7284 | CAGGCTTCCAAGTACCCATCTCGTCCAGGGTTCTCGCACCCACCGCGACGCCCGG | 7225 |
| QY | 603 | uLeuAlaSerMetGluLeuAsnGluSerSerArgTTPThrGluGluGluMetGluThrAl | 623 |
| DB | 7224 | ATGCCGG-----CACGCCTCGATGAAGCGGT | 7198 |
| QY | 623 | aLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySe | 643 |
| DB | 7197 | G-----TGCGCGCG-- | 7188 |
| QY | 643 | rLysThrValSerGlnCysLysAsnPheTyrPheAsnTyr-LysLysArgGlnAsnLeuAs | 663 |
| DB | 7187 | -----CGACACAGCCCTGT----- | 7173 |
| QY | 663 | pGluIleLeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAlaArgIysLys | 683 |
| DB | 7172 | -----GTCCGCTCCCGAAGCGGACGACCTGGCGCGAGTTGCG | 7135 |
| QY | 683 | sLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGl | 703 |
| DB | 7134 | GTACCAGTAGTCGCGCTCGAGTTCGGTCCGCTCGATCGGTTCCGCGGTGACGGTGGAGTA | 7075 |
| QY | 703 | uGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAl | 723 |
| DB | 7074 | -GAAGGCACGTCCCGGTCTCCGCGGATGGGGCGAGGA-GTTCGACGAGCC----- | 7022 |
| QY | 723 | aLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAs | 743 |
| DB | 7021 | -----GCCCTCGATCCCTCGACCTGCG-GGGGAGTGGCACGCCCGACGTCGACATGAG | 6970 |
| QY | 743 | nAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGl | 763 |
| DB | 6969 | CCTCGGGCCCGGTACGTCTTCTGTGACAGCTCTCTCCACGAGGCGGTCCGCGTCCCGG | 6910 |
| QY | 763 | yGln-AsnGlyProLysPro----- | 769 |
| DB | 6909 | CTCCCGGAGACCAAGCGCGGAAACGGCCCGTTCAGCGCGCGATGACCGACGCGTCCGC | 6850 |
| QY | 770 | -----ProAlaThrLeuGlyAlaAspGlyProProGlyProp | 783 |
| DB | 6849 | CCATGTCCAAAGACCGCGCTCCAGCTTCTCGACCGCGAGACCGACCGATGTATCGCCCC | 6790 |

[illegible]

PD 31-AUG-2000.
 XX
 PF 23-FEB-2000; 2000WO-JP001041.
 XX
 PR 24-FEB-1999; 99JP-00046961.
 XX
 PA (KITA) KITASATO INST.
 XX
 PI Omura S, Ikeda H;
 XX
 DR WPI; 2000-565458/52.
 XX
 DR P-PSDB; AAB23751, AAB23752.
 XX
 PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
 PT it for the production of avermectin and its derivatives for drug and
 PT agrochemical use.
 XX
 PT
 XX
 PS Claim 2; Page 134-203; 314pp; Japanese.
 XX
 CC The present sequence represents DNA which encodes avermectin aglycon
 CC synthase proteins. Also described are: (1) polypeptides encoded by all or
 CC part of the DNA; (2) expression vectors containing the DNA; (3) host
 CC cells transformed by the vectors; (4) preparation of the polypeptides by
 CC culture of the transformants; (5) preparation of avermectin aglycon or
 CC its derivatives by culture of transformed avermectin-producing
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
 CC containing sense or antisense sequences from the avermectin aglycon
 CC synthase DNA. The enzymes are useful for the production of modified forms
 CC of avermectin and of the intermediates in its biosynthesis, for use as
 CC drugs, veterinary drugs and agrochemicals
 XX
 SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9.14e-11 Length: 31422
 Score: 639.50 Matches: 633
 Percent Similarity: 31.27% Conservative: 286
 Best Local Similarity: 21.54% Mismatches: 1158
 Query Match: 4.04% Indels: 875
 DB: 3 Gaps: 122

US-09-522-753-5 (1-2517) x AA92302 (1-31422)

QY 9 AlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProVal 28
 DB 8890 GCCGTAGTGGCGAGCAACGCTCGCCTCGATCGATCCACCAACGTCGTACCCGTC 8831
 QY 29 GlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp 48
 DB 8830 CGTGCCCTCCACCAT-----CCACATCCGCCACCGACA 8795
 QY 49 TyrAlaSerHisLeuSerProGlySer-----IleIleGlnProGlnArgArgPro 66
 DB 8794 ACCCGCACAGCCCAACGCTCGGAATCACCCTCGCTCGAGCGGACCAATTCGGCGCC 8735
 QY 67 SerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgPro 86
 DB 8734 TCAACCATTCGACGACCGCTCTGATTCCACCGCACTCCCGCGCACCCGCCCAAAACC 8675
 QY 87 GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
 DB 8674 GATGACCACACGCTTACGCTCGAGACGCGCTCCACCAACAGCACACCCACCTCGG 8615
 QY 107 ArgProArgLeuGluLeuProAspProLeuLeuArgProSerProLeuLeuAlaThr 126
 DB 8614 CCCAGCGA-----CCCATCGGCCCGACCCCTAGCTTCGCTTCGAC 8573
 QY 127 GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu 146
 DB 8572 GCGCTCGGCGACAGACCCGCTCGCGCAG-----ACTCCACAAACGCAC 8525
 QY 147 ProValSerProProSerProProHisThrAspProGluLeuLeuValProProArg 166

DB 8524 CCGGCGCTCG---ACATCACCGTCCACACCCCGCCCAACGCCAGCAACTCCCCCG--- 8471
 QY LeuSerLys-GluGluLeuIleGlnAsnMetAspArgValAspArgValAspArgLuleThrMetVa 186
 DB 8470 -----ACCTCAACGCTGACACCGCCAGATGACGCGCCACCCACGACGACGACACA----- 8422
 QY 186 lGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnGlnLeuGluGluGlu---AlaAl 205
 DB 8421 -----CGCCGTATCCACCGTCCACCGCGGACCCCTCGAAGCCGAGGTGTAGG 8375
 QY 205 aLysProProGluProGlu----- 211
 DB 8374 AAGCGCCCGGAGACGACGCTTTGGAGAGCGCGGTGAGCGGTACCCCTCTGTGTCT 8315
 QY 211 ----- 211
 DB 8314 GGGTGGCGGCGGACGAGCTCGGCGTAGTCTCTGTCGAGAGCGCGGCGGAGACACCGG 8255
 QY 212 -----LysProValSerProProProlleGl 220
 DB 8254 TCGTGACCCGTGACGCGGTGGGGTGCATGCCCGCGCTCCAAACGCTCCAGGACA 8195
 QY 220 uSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAl 240
 DB 8194 CTCCAGCATCAACCGCTCTCGGATC-----CATGCCAACGCTCAGCG 8147
 QY 240 aAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu----- 254
 DB 8146 GACT-----GATCCCGAANAACCCGCACTCGCGCGCACCTCGCAAAA 8099
 QY 255 -ProLeuTyrAsnGlnProSerAspThrArg-----GlnTyrHisGluAenIl 270
 DB 8098 ACCCACACACCGGTGTAGGACGTATCCCGCGCGCGCTCCGGATCATAGAAAGCCT 8039
 QY 270 eLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgAsnHisAl 290
 DB 8038 CCAGTCCCAACCCCGGTGACCGGAAATC----- 8008
 QY 290 aArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGl 310
 DB 8007 -----CCCCACCGCATCCGACCCGCGCGATCACTCCAGAAATCTCTCCG 7961
 QY 310 uLysLysValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysValAr 330
 DB 7960 CCGACTC-----CACACCCCGCGGAAACCGCACGCCA-TCC----- 7925
 QY 330 gGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnArgMe 350
 DB 7924 -----CCACAATGGCAATCGGCTCGTCGACATCCACACCGCGGT 7887
 QY 350 tGlnSerArgValGlyGln-----ArgGlySerGlyLeuSerMetSerAl 365
 DB 7886 CCGCGGCGCGAAGGCGNAGGCGAAGGCGAGTCCGCTGGTGGCTTCGCGCTCAGC 7827
 QY 365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAenLe 385
 DB 7826 TGCTCGTGGATG-TGTCCCGAGTGCACCGCGCGGAGTGTGCGAACACCGAGGTCTCT 7768
 QY 385 uGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGl 405
 DB 7767 CGGAAGCGC---AGGCCGTGGCGGTGTG----- 7740
 QY 405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425
 DB 7740 ----- 7740
 QY 425 pArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPheArgGluLysPheMe 445
 DB 7739 -AGCGGTTCGTAGTTCGACGCGGTGAGGAGTTCGAAACCGAGGTTCGCGGAAG----- 7686
 QY 445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeu----- 458
 DB 7685 -----GCGCGTTCGGGACACCCGATCCGCGGTGCGGTGTCTCAAGAGCGCGCG 7636

| | | | |
|----|-------|--|-------|
| Qy | 1999 | -AlaLysAsnLeuAlaProHisHisAlaSerProAspProAlaPro----- | 2014 |
| Db | 40277 | CGCGACGGCGTGGAGCCCATTCACAGAGCGCTGGGACGCGGTGCTCTACGACCCC | 40336 |
| Qy | 2014 | ----- | 2014 |
| Db | 40337 | GACCCGACCGCAGCCCAAGAGCTACCTCGCGCATCGCGGATGCTCGACCATCGAC | 40396 |
| Qy | 2015 | -----ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr | 2030 |
| Db | 40397 | CTCTTCGACCGCTCTCTTCGGCATCAGCCCCCGGAGGCCAAACACC---TCGACCCC | 40453 |
| Qy | 2030 | opheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrrHisGlySerSerTyrrSerPr | 2050 |
| Db | 40454 | CAGCACGGCGCTGCTCTCGAAATCTGCTGCTGGCTGGCCCTCGAGGACGGCGGATCGTCCC | 40513 |
| Qy | 2050 | oGluGlyValGluProValSerProValSerSer----- | 2061 |
| Db | 40514 | ACCTCCCTCAAGACTCCCTCACCGCGCTTCCTGTCGGCATCTGCGCGCGGAATACGCG | 40573 |
| Qy | 2062 | -----proSerLeuThrHisAspLysGlyLeuProLysHis | 2073 |
| Db | 40574 | ATGCAAGAGCGGAGCTCGGAAGGTTCCGAGGTTTACTTCATCCAAGGCATCTCCG--- | 40628 |
| Qy | 2073 | sLeuGluGluLeuAspLysSerHisLeuGluGlyLeuLeuArgProLysGlnProGly-- | 2092 |
| Db | 40629 | -----CGTCTTTGGCGCGGGGGCG | 40648 |
| Qy | 2093 | ---ProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGl | 2111 |
| Db | 40649 | TTGGCCTATACGCTCGGGCTCCAGGGCGCG---GATCTTCTGTCGACACCGCC | 40699 |
| Qy | 2111 | uSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr | 2131 |
| Db | 40700 | TGCTTCCTCTCGCTCGCTCCC---TCCACCTCGCCTGCCAAGCCCTCGACAGGGC | 40753 |
| Qy | 2131 | gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrrThrArgHisHi | 2151 |
| Db | 40754 | GAGTGCACCC----- | 40763 |
| Qy | 2151 | sProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrrSerPhe-----Pr | 2167 |
| Db | 40764 | -----TCGCGCTTCGCGCGGGCGTCTGCTCATGCTCTCCCCCAGACC | 40807 |
| Qy | 2167 | oGlyAlaSerCysProValLeuAspLeuArgProProSerAspLeuTyrrLeuProPr | 2187 |
| Db | 40808 | TTCCGTATCTCTTCCTGCTGCGCGCTTGGCGCCCGCAGCGGC-----GCTCCAAG | 40858 |
| Qy | 2187 | oProAspHisGlyAlaProAla-----ArgGlySerProHisSerGl | 2201 |
| Db | 40859 | ACCTTCTCGGACACCGCGGCTACGGACGGGAGAGGGGTCTGCTCTTCCCTC | 40918 |
| Qy | 2201 | uGlyGlyLysArgSerProGluProLeuLysThrSerValLeuGlyGlyGluAspGl | 2221 |
| Db | 40919 | GAGCGGATCGCGCAGCGCCTCGCCCGGAGACACCGCG-----TCCTC | 40960 |
| Qy | 2221 | ylleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVa | 2241 |
| Db | 40961 | GTCCTCTGTCGGCGCACCGCATCAACACGACG-----GCGCGTCGAGCGGT | 41008 |
| Qy | 2241 | lTyrrProLeuLeuTyrrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe | 2261 |
| Db | 41009 | ATCACCG-----CCCCCAACGGCACCTCCACGACAGA-----AGGTC | 41044 |
| Qy | 2261 | rProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAl | 2281 |
| Db | 41045 | CTCCGGGGCGGCTCCACGACGCGCGCATCACCCCCCGCAGCTCGACGTCGTCGAGTGC | 41104 |
| Qy | 2281 | aMetValLysSerLysLysGlnGluIleAsn----- | 2291 |
| Db | 41105 | CATGCACGGCACCTCTCGTCGGGAGACCCCATCGAGGTGCAAGCCTGGCGCGCTCTAC | 41164 |
| Qy | 2292 | -----LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrrAs | 2305 |

| | | | | |
|------------|--|--|--|-------|
| Db | 41165 | GC | CGACGGCAGACCGCTGAAAGACCTCTCCTTCGGCGCGCTCAAGACCAACATCGGC | 41224 |
| Qy | 2305 | nile | SerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMe | 2325 |
| Db | 41225 | CATCTCGAGCGC | -----CCTCGCGCCTC-GCGGCGCTCGCCAAGAT | 41265 |
| Qy | 2325 | tThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaI | 2345 | |
| Db | 41266 | GGTC | -----GGCTCGCTCCGACGAGCGCCTGCCCCACCCCTCCACGCGAC | 41313 |
| Qy | 2345 | eileArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlu-GluserProProLeuSerA | 2365 | |
| Db | 41314 | CCACGCAATCCCTCATC | -----GAGTGGAGCGCTCGCCATCGACGTCGT | 41361 |
| Qy | 2365 | laAsnAlaPheAsnProLeuAsnAlaSerLeuProAlaAlaMetProIleThrA | 2385 | |
| Db | 41362 | CG | -----ATACCCGAGCGCTTGCCCGCGCACGAAGATGCGATGCCCGCCGCGC | 41412 |
| Qy | 2385 | laAlaaspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLys | 2402 | |
| Db | 41413 | CGGCATCTCGCTTCGGATTCTCGGCGACCAACGCCCCAGCTCATCTCCGAGAGGCTCC | 41472 | |
| Qy | 2403 | -----AlaLysValSerGlyAArgProSerSerArgLysAlaLysSer | 2419 | |
| Db | 41473 | CGCGGCTTCGGCGCGACCGCCACCTCACAGCGCGCTCGAAGCGCTCCCGCGGC | 41532 | |
| Qy | 2419 | roGlyLeuAlaSer | -----GlyAspArgProProSerValSerSerValHisSerG | 2436 |
| Db | 41533 | GTGCGCGCTGCTCCTGTCGGCGAGGAGCGAGCGCGC | -----TCCGCGCCCGAGCGGAA | 41586 |
| Qy | 2436 | luGlyAspCysAsnArgArgThrProLeuThr | -----AsnArgValTrpGluAspA | 2453 |
| Db | 41587 | CGGCTCCGGCACCACTCGTCGCCACGACGACCTCACCTCCGGGATGTGGCTATTC | 41646 | |
| Qy | 2453 | rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA | 2473 | |
| Db | 41647 | GC | ----- | 41648 |
| Qy | 2473 | laGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaG | 2493 | |
| Db | 41649 | -----TGGCACCAACCGCGCCACTTCGAGCACCGCGCGCTCTCGT | 41691 | |
| Qy | 2493 | lyProHisHisAlaTrpAspGluGluProLys | 2503 | |
| Db | 41692 | AGCCACACCGCGAGGAGTCTCTCTCCGCGC | 41723 | |
| RESULT 57 | | | | |
| AAA92302/c | | | | |
| ID | AAA92302 standard; DNA; 31422 BP. | | | |
| XX | AAA92302; | | | |
| XX | 10-JAN-2001 (first entry) | | | |
| DE | S. avermitilis avermitilis aglycon synthase DNA aveAll SEQ ID NO:2. | | | |
| KW | Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis; | | | |
| KW | multifunctional enzyme; polyketide; avermectin; veterinary drug; | | | |
| KW | agrochemical; ds. | | | |
| OS | Streptomyces avermitilis. | | | |
| XX | Key | | | |
| FH | Location/Qualifiers | | | |
| FT | 1..14646 | | | |
| FT | /*tag= a | | | |
| FT | /note= "avermectin aglycon synthase protein" | | | |
| FT | 14824..31422 | | | |
| FT | /*tag= b | | | |
| FT | /note= "avermectin aglycon synthase protein" | | | |
| XX | WO200050605-A1. | | | |
| XX | | | | |

Db 38173 CGCCTCCACGGCAGCCCGCCCGCGGTCAACACCTTGTCTCCGCTCG-----AGGG 38223
QY 1475 erLeulleGlySerProGlyArg----- 1482
Db 38224 GCGTTCCTGCGAGGCCATCGAGAGCGGAATATCGAGCGCTCAGCGGCCAGCTCCACGT 38283
QY 1483 -----ThrPheProProValHisProLeuA 1491
Db 38284 GGACGGCGACGAGCGCGCCGCTTGCCTCTCTTCCACACCTCGCGAGCTTCG 38343
QY 1491 spValMetAlaAspAlaArgAlaLeu-GluArgAlaCysTyrGluGluSerLeuYsSer 1510
Db 38344 CCACGAGCGGCAAGAGCAGGCGACCGTGCAGCGCTGCGCTACCGATCAGCTGGAGAGC 38403
QY 1511 -----ArgProGly-----ThrAlaSerSerSerGly 1519
Db 38404 TCTGACCACCGCCACACCGCCCGACCTGGCGGACCTGGCTCTCTGCTGCGCGC 38463
QY 1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyysPro-----Arg 1537
Db 38464 CGCTCTGGACGACGAGCGCTCCCTCCGCGCTCACCGAGCGCTCGCCGCGCGCGC 38523
QY 1538 GlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly 1557
Db 38524 GCGGTCCTCGCGCTGCGCTGAGCCAGGCCACCTGGA-----CCGCGAGGC 38571
QY 1558 SerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSer 1577
Db 38572 TCT-----CGCCGACACCTGCGCCAGGCTTGGCGGAGACCGCGCGCTCGCG 38622
QY 1578 LyAlaSerGlnAspArgLysLeuThrSerThrProArgGluLeAlaLysSerProHis 1597
Db 38623 CGTGCTCTCGCTCTCGCCCTCGACGAAAGTCCCTCGCGACCATCGCGCGCTGCGCGC 38682
QY 1598 SerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGly 1617
Db 38683 GGGACTCGCTCTCGCTCACCTCTCGTCCA---AGCCCTCGCGACATCGCCCTCGACGC 38739
QY 1618 ValSerGlyValAspLeuTyr-----ArgSerHisIle 1628
Db 38740 GCCCTT---GTGGCTCTTACCCCGCGCGCGTCTCGTCCGACACTCCGACCCCATCGC 38796
QY 1629 Pro-----LeuAlaPheAspProThrSerIleProArgGlyIlePro 1642
Db 38797 CCATCCGACGACGCGATGACCTGGGCGCTCGCGCGCTCGTCTCGCGCTCGAGCA---CCC 38853
QY 1643 LeuAspAlaAlaAlaTyrTyrLeuProArgHis----- 1654
Db 38854 CGAGCGCTGGGGAGGGCTCGTCCGACGTCCGCGCGACGATCGACGCGAGCGCCGTGGGCGC 38913
QY 1655 LeuAlaProAsnProThr-----TyrProHisLeuTyrProPro---Tyr 1668
Db 38914 CTGTCTCCCGTCTCTCGCCCTCGCGACAGATGAGACCACTCGCTCTCGCCCGCGCGC 38973
QY 1669 LeulleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
Db 38974 GTTCTAGCTCGCGCTCGCTCGCGCTCGCGCTCGCGCGCGCGCGC-----CGCAGC 39027
QY 1689 TyrIleThrSerGlnGln-----MetHisAsnThrAlaThrAlaMetAlaGlnArg 1706
Db 39028 TACTTCAGCCCGGAGGCCCTCTCTCATCACCGGAG-GACCGGCGCGCTGGCGCTC 39086
QY 1707 AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAla 1726
Db 39087 ACGTGCGCCGATGGCTCGCTCGAGAGCGCGACAGCACCTCGTCTCATCAGCGCGCGAG 39146
QY 1727 GlyProArgGlyIleLeuAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
Db 39147 GGGCCAGCGCGAGCGCGCTCGAGCTCCACGCGCGAGCTCACGCGCCCTGGCGCGCGC 39206
QY 1747 ThrProGly-----ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPro 1762
Db 39207 TCACCTTCGCGCGCTGTGATGTCCGCGACAGGAGCGCTGTGCGCACGCTTCTCGAGCAGC 39266

QY 1763 ThrAlaProGln-----ProPheSerSerArgHisSerSerSer----- 1775
Db 39267 TCGACCGCGAAGGGTTCGAGGTCGCGCGCTGTTCACGCGGCGGATCGCGCGCCACG 39326
QY 1776 -----ProLeuSerProGlyGlyProThr----- 1783
Db 39327 CTCGCTCGCGCCACCTCTCTCATGGAGCTCCCGACGTTGTCTCTGCCAAGTCTCTAG 39386
QY 1784 -----HisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db 39387 GCGACGGGAACCTCCACGACCTGCTCGTCTCGACCCCTCGACGCTTCTCTTCTTCT 39446
QY 1796 Arg-GluArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLe 1815
Db 39447 CGTCCATCGCAGCGCTGGGGCGCGGACACAAGCGGATACGCGCGCGGAACGCCT 39506
QY 1815 uThrSerThrThrValGluHisAlaProIleTrp----- 1827
Db 39507 TCTCGACGCTTGGCGGACCGAGCGCGAGTCTTGACAGCGGACACGTCGCTGGTGT 39566
QY 1828 -----ArgProGlyThrGluGlnSer 1835
Db 39567 GGGCGCTGGGGCGCGGGTGTATTATTCACGGGGCCCTTGGCAGCCAGCTGGAGC 39626
QY 1835 rGlySerSerGlySerSerGly-----GlyGlyGlyGlySerSerSerArgProAl 1852
Db 39627 AACGTCGTCTGTCGCGATGGCCCTTTCGTCGCTGGCGGCGCTCGCGCAAGCCCTGG 39686
QY 1852 aser-----HisSerHisAlaHisGlnHis----- 1860
Db 39687 AGCAGCAGACACACCGTCACCGTCGCGACATCGACTGGGCGCGCTTTCGCCCTTCGA 39746
QY 1861 -----SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLe 1878
Db 39747 TCAGCGCTCGCTCGCGCGCTCTCGCGACTTTCGCGAGCGCGC----- 39795
QY 1878 uHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLe 1898
Db 39796 -----GCCCTCGAAGAC----- 39807
QY 1898 uArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHi 1918
Db 39808 -AGAGAAGCGCGCTCTCTCTCCGACACGCGCCCGCGCCCGCCCTCTCGCAAGCTCC 39866
QY 1918 sCysProLeuGlyGlyThrLeu-----AspG1 1927
Db 39867 GGAGCGCTCGAGAGCGGAGCAGCTCGTCTCGCGCGCTGGTGTGCGAGGAGCGG 39926
QY 1927 YValTyrProThrLeuMetGluProValLeuLeuLeuProLysGluAlaPro---ArgValAl 1946
Db 39927 CCCTCGCTCGCGCCACGAGGCG-----CTTCGAGCTCGACCCCGACAAAGCTTCT 39980
QY 1946 aArgProGluArgProArgAlaAsp-----ThrGlyHisAlaPheLeuAlaLy 1962
Db 39981 TCACCT---CGTCTCGATTCGATCATGACCTCGAGCTTCTGTCGCGCTTCAAC-AG 40036
QY 1962 sProProAlaArgSerGlyLeuGluProAlaSerSer----- 1974
Db 40037 GCCACCGGATCAAGCTCCCGGCCACCTCGCTTCGACCATCTCTCTCATCGCGTC 40096
QY 1975 -----ProSer---LysGlySerGluProArg-- 1982
Db 40097 GCGCTCTTCATGCGGACTCGCTCGCCCGCGCGCTCGGACAGGCTCTCGCGCGAGCG 40156
QY 1983 -----ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
Db 40157 ACGCCCGCGCTCCGCGCGCGCTTCGAGCGAGAGCCCATCGCTCGTCGCGATGGCC 40216
QY 1998 o----- 1998
Db 40217 CTGCGCTGCGCGCGCGCTCGCGGATGTGACGCTCTTTGGGAGTTCCTCCACCAAGG 40276

Qy 854 pThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyPr 874
Db 36194 AAGACCTTCGACCAACGCGACGGCTACGACGCGGAGGCGCGTCTGCTGCC 36253
Qy 874 oAlaLysGlyLysAspAlaGluAlaGluAlaThrAlaGluGlyAlaLeuLysAlaG1 894
Db 36254 CTCGAGCGGTTCGCGGACGCGCTCGCGGAGGACACGCGCTCTCGCGCTCTCGCGGCG 36313
Qy 894 uLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAs 914
Db 36314 ACCGCCAAGAACATGACGGCGCTCGAGCGGATCACCGGCCCAATGGCACCTCCAC 36373
Qy 914 pSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLy 934
Db 36374 CAGAGGTCTCTCGCGCGCGCTCCACGACGCCCATATCGGCCCTCGCGAGTCGACGTC 36433
Qy 934 sAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAs 954
Db 36434 GTCGAATGCCATGCGACGCGCACCTCTTGGGAGACCCCATCG-----AGGTGCAA 36484
Qy 954 nAla---SerProGlnLysProLeu-----AspLeuLysGlnLeuLysGlnArgAlaA1 971
Db 36485 GCCCTGGCGCGGTCTACGCGGATGGGAGACCCCGCTGAAAGCCTCTCTCTCTCGGCGCA 36544
Qy 971 aAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaA1aPr 991
Db 36545 CTCGAACCAACATGGCCATCTCGAG-----CCGCTTCGCGCTCGCGGCGTCTGCC 36598
Qy 991 oThrLysProAlaPro-----ProAlaProProPro-----ProGlnAs 1004
Db 36599 AAGATCGTCGCTCCCTCGGCATGAGCCCTCGCCCGCCACCTCCACACGCCCGCGC 36658
Qy 1004 nLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer-----ProArgGlyLy 1022
Db 36659 AATC---CCCTGTAGTGGATGGCTCGCTCCCATCGACGTCGTCGATGCCACGAGGCG 36715
Qy 1022 sSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe 1042
Db 36716 TGGG-----CCCGCCACGAGATGGCAGTCCCGCGCGCGCGGCTCTCC 36760
Qy 1042 uProGlyAspProProCysThrThrSerGlyLeuProPhe----- 1055
Db 36761 GCCTTCGAGCTCTCGGCACCAACGCGCCACCTTATCTCGAAGAGGTCTCCGCGATCCCG 36820
Qy 1056 ---ProValProArgGluValLysLysAlaSerProHisAlaProAspProSerAl 1074
Db 36821 CAGCGGAGCCACCGCGGACAGCTCGCTCGGAGC---CGCTTCGCGAGCTGCGCC 36877
Qy 1074 aPheSerTyraLaPro-----ProGlyHisProLeuProLeuGlyLeuHisAs 1090
Db 36878 GTGCTCTCTCGGCAGGAGCGCGCGCTGCGCGCCAGGCCAGGCCAGAGCTCCGCGAC 36937
Qy 1090 pThrAlaArgProValLeuProArgPro-----ProThrIleSerAsnProPr 1106
Db 36938 CACCTCTCTCGGCCACGACCTCGCCCTGCGCGATGTAGCTACTCTGCTCGCCACCAACC 36997
Qy 1106 o-----ProLeuIleSerSerAlaLysHisProSerValLeuG1 1119
Db 36998 CGGGTACTCTCGACCGACCGTGGCTCTGCTGTCGACGACCGCGAAGAGCTCTCTCC 37057
Qy 1119 uArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrsSerG1 1139
Db 37058 CGCTCGATTCTGCTCGCCCGAGGAAGGCCCGCC-----CGAGCACCGTCTGCGAA 37108
Qy 1139 uHiAlaLysAlaPro-----ValGlyPro-valThrMetGlyL 1152
Db 37109 CGAAGCGGAAGCCACGGAAGTCTGCTCTCTCTCTTCTTCTGGGCAAGGCTCGCAGTGGAA 37168
Qy 1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGly---ValLysGlnGluG 1171
Db 37169 GGGATGGCCCT 37228
Qy 1171 InLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG 1191

Db 37229 GAGCGCGCTCTCGCGCCACCGTGGAGTGTCTGCTCGCGTCTCCGCG----- 37281
Qy 1191 luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG 1211
Db 37282 -----CGAGGA-----GGCGCGCCCGCTCGACCGCGTTCGAGTG 37318
Qy 1211 lyIleProSerThrArgValProSerAspSerAlaIleThrTyArgGlySerIleThrH 1231
Db 37319 GT-----CCAGCCCGCGCTCTCGATGATGTCTGCTGCGCGCCCTGTGCGGTCC 37372
Qy 1231 isGlyThrProAlaAspValLeuTyLysGlyThrIleThrArgIleIleGlyGluAsps 1251
Db 37373 ATGGCGCTCGAGCC-----CGACGCGGTGGTTCGCGCAT 37405
Qy 1251 erProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrg 1271
Db 37406 AGCCAGGCGAGATCGCGCGCG-----CTGTGTGGCGGCGGCTGCTCGCTC 37453
Qy 1271 luGlyLysLysGlyHisValLeuSerTyrgLysGlyMetSerValThrGlnCysSerL 1291
Db 37454 GAGGACGC-----TGCCAAGCTGTGGCTGCGCAGCGGTGCGCTC 37495
Qy 1291 ys-----GluAspGlyArgSerSers 1298
Db 37496 GTGGAGCTCGCGGCCAGGGGCGCATGGCGCGGTGGAGTGGCGGAGCGCGGATCGCA 37555
Qy 1298 erGlyProHisGluThrAlaAlaProLysArgThrTyArgMetMetGluGlyArgV 1318
Db 37556 CGGCGCTCCA-----GGCTATGGCGATCGGCTC 37585
Qy 1318 alGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProG 1338
Db 37586 TCCATCGGGCGATCAACAGCCCTCGTTTCACGACGATCTCCGCGGAGCCCGCTCGCGTC 37645
Qy 1338 luArgHisSerProHis-----HisLeuLysGluGlnHisIleArg- 1352
Db 37646 GCGCCCTGCTCGCGCATTCGAGTTCGAGGGGCTCTTCGCCCTCAAGCTGAGTTAGCAC 37705
Qy 1353 -----GlySerIle---ThrGlnGlyIleProArgSerTyrgValG 1365
Db 37706 TTGCTCTCCACTCCGCGGAGTGTAGTGTCTACTTCTACCGGTGAGCGCGCGCGCATC 37765
Qy 1365 luAlaGln-----GluAspTyrgLeuArgA 1373
Db 37766 TGGCTCGAGCGCGCTCGACGCGGTCCCGTCTACTTCTACCGGTGAGCGCGCGCGCATC 37825
Qy 1373 rgGluAlaLysLeuLeuLysArgGlyThrProProProPro----- 1387
Db 37826 GACGGGAGAGCTCGACGCGCGCTACTGTGTACCGGAACCTTCGCGGACCGCGTCCGCTTC 37885
Qy 1388 -----ProProSerArg----- 1391
Db 37886 GCAGACGCTGTGCAAGCGCTCTTCGCGGAGAACATCGTCTCTTCGTGGAGTGGAGCCCC 37945
Qy 1392 -----AspLeuThrGluAlaTyLysThrGlnAlaLeuGlyProLeu----- 1405
Db 37946 AGTCTGTGTGACCTTGGCTTCGACGAGCTCTCGAAGCGTTCGAGCGCTCGCGCGCG 38005
Qy 1406 -----LysLeuLys---ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
Db 38006 GTGTGTGCTCTCTGTGTGGAGCGACGAAGGGGATCTACGGCGCTTCCTCTGCTCTCTCTCC 38065
Qy 1423 rgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaP 1443
Db 38066 GAGTCTACGTCAACGGCTTCGCCCTGGATGGAGACGACATCT-----GCC 38113
Qy 1443 roArg-----ProLeuLysGluGlySerIleThrGlnGlyT 1455
Db 38114 CCGGGAAGCGGTGCGCTGCCACCTACCCCTTCC-AGCGCGAGCGCTTCTGCTCGA 38172
Qy 1455 hrProLeuLysTyArgThrGlyAlaSerThrGlySerLysLysHisAspValArgS 1475

Db 34281 AAGACCAAGCTCGCGCTCGCCC-----GGCGCGCTCTACGACGCC 34322
Qy 255 oLeuTyAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275
Db 34323 GCTTCGTCGCGCGCGCTCGG-----CGATCGCG 34352
Qy 275 aMetArgLysLeuLeuLeuTyrPheTysArgAsnHisAlaArgLysGlnTrpLy 295
Db 34353 CTGCGCTCGCGGCTT-----CATGCCCGAGCACCACCTCATCACCGGTGTA 34403
Qy 295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLys-----Ly 312
Db 34404 CCGCGCGCATTCGCGCTACGCTCGCCGATGGCTCGCTCGAAAGCGCGTACACCTCG 34463
Qy 312 sValGluArgIleGluAsnProArgArgAlaLysGluSerLysValArgGluTy 332
Db 34464 TCCTCATCAGCCAGAGGGCGCCAGCCGAGGGCGGTGGAGCTCCAGCGGAGTCA 34523
Qy 332 rTyrGluLysGlnPheProGluIleArgLysGlnArg---GluLeuGlnGluArgMetG 351
Db 34524 CCGCCCTCGCGCGCGCTCACCTTCGCGCTCGCGTGCATGTCGCGACAGGAGCGCTGTCG 34583
Qy 351 n-----SerArgValGlyGlnArg-GlySerGlyLeuSerMetSer----- 364
Db 34584 CCACGCTTCTCGACGAGCTCGACGCGCGA---GGCCACAGGTGAGCGCGCTGTTCCAG 34640
Qy 365 -----AlaAlaArgSerGluHisGluValSerGluI 375
Db 34641 CGGGCGCATCGAGCCCGCTCGCGCTCGCGCCACCTCATGGAGGATCTGCCGAGG 34700
Qy 375 leIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValI 395
Db 34701 TTGTCTCCGCAAGGTACAAGGT-----GCAAGACACCTCCACGACCTGCTCGGCT 34751
Qy 395 leProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyL 415
Db 34752 CTCGACCCCTC----- 34762
Qy 415 euMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluG 435
Db 34763 -----GACGCTTTGTTCTCTCTCTCGCGCGCGGTC-----GTCTGGGGCGCG 34808
Qy 435 lnGluLys-----GluThrPheArgGluLysPheMetGlnHisProL 449
Db 34809 GACAAACAAGCGGTATGCGCGTGCAGAACCTTCTCGATCGCTCGCGCGCGAGCGGC 34868
Qy 449 ysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuT 469
Db 34869 GCAGCCTTGGGCTGACGGCGACA-----TCGGTGGCCT 34901
Qy 469 yRTyTyrLeuThrLysLysAsnGluAsnTyrLysSer----- 481
Db 34902 GGGG-CGTGTGGGCGCGCGGCGATGGTACCGGGCTCTCGGAGCCAGCTAGACCAA 34960
Qy 482 -----LeuValArgArgSerTyrArgArgGlyLysSerG 494
Db 34961 CGCGGTCTGTCGCGATGGCCCTCTGCTGCGCGTGGCGAGCTCGCGTGGCGTGGAG 35020
Qy 494 lnGlnGlnGlnGlnGlnGlnGln-----G 505
Db 35021 CACGACGAGACCACTTCACCGTGGCGGACATCGACTGGCGCGCTTTGGCGCTTCGTTC 35080
Qy 505 lnGlnGlnGlnGlnGlnProMetPro-ArgSerSerGlnGluLysAspGluLysGlu 524
Db 35081 AGCGCGCTGCTCCCGCGCTCTCTGCGGATTTGCGGAGGGCGCGCGCTCTCGAA 35140
Qy 525 LysGluLysGluAlaGluLysGluGluGlu-LysProGluValGluAsnAspLysGluAs 544
Db 35141 GCCAGCGCGATGGCTCTCGAGCAAGAGCGGGCCACAGCGCTCTCTCG----- 35189
Qy 544 pLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaVa 564
Db 35189 ----- 35199

Qy 564 lAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgLysGlyArgIleThr-- 583
Db 35190 -----ACAAGCTCCGAACCGCTCGGAGAGCGAGACAGATCCACCTG 35230
Qy 584 -----ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIle---ThrProGlnGl 600
Db 35231 CTCTCTCGTGTGGCGCCACGAAGCGCGCTCTGCTGGGCCATACACGAGCTCCAG 35290
Qy 600 nSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluMe 620
Db 35291 GTCGACCCCC-----ACAGGGCTTCATGGACCTTCGGCTCGAT 35329
Qy 620 tGluThrAlaLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMe 640
Db 35330 TCCTCATGACCGCTCGAGCTTCGTC-----GGCGCTTGCAGCAGGCC 35371
Qy 640 tValGlySerLysThrValSerGlnCysLysAsnPheTyr-PheAsnTyrLysLysArg 660
Db 35372 ACCGGCATCAAGCTCCGGCCACCTCGCTTCGACCATCTCTCTCATCGCGTCGCG 35431
Qy 660 lnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaA 680
Db 35432 CTCTTCTGGCGA---CTGCTCGCCACCGCTCGCGCGAGGCTCTCGTCGAGCGC 35488
Qy 680 rArgLysLysLys-LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Db 35489 GACGCGCGCGCTCCCGCGCTTCGCTCGCGAGCGACGAG-----CCCATCGCC 35539
Qy 700 valGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGlu 719
Db 35540 ATGCTCGGATGCGCTCTCGCTTCGCGCGCGGATCGCGCATGTGCGACGCTCTTTGGAG 35599
Qy 720 -----GluAlaGluAlaLeu---HisAla-SerGlyAsnGluVa 731
Db 35600 TTCTCGCCCAAGGACGCGCGCTGCGACCCATTCCTCCCATGCGCATGGATGCGGCT 35659
Qy 731 lProArg-----GlyGluCysSerGlyProAlaThrVa 742
Db 35660 GCCTCTACGACCCCGACCCGCGAGCCAGCCAGAGTACGTCCGCGCATGCCGCGCATG 35719
Qy 742 lnAsnAsnSerSerAspThrGluSerIleProSer-----ProHi 755
Db 35720 CTCGACCGAGTGCACCTCTCGATCTCTGCTCTTTGGCATCAGCCCTCGCGAGGCCAAA 35779
Qy 755 sThrGluAlaAlaLysAspThr--GlyGlnAsnGlyProLysProProAlaThrLeuGl 774
Db 35780 TACTCGACCCCGACGACCGCTGCTCTCGAATCTCGCTGGCTGGCGCTTCGAGGACGCC 35839
Qy 774 yAlaAspGlyProProGly-----ProProThr-- 784
Db 35840 GGATCGTCCCTCCACCTCAAGGATTCCTCCACCGCGCTCTTCGTGCGCATCGCGGCC 35899
Qy 785 -----ProProArgThrSerArgAlaProIleGluPr 796
Db 35900 AGCGAATACACACTGCGAAACACGAGCTCGAAGAGGTGGAAGGTATGCGCTCCAAAGC 35959
Qy 796 oThrProAlaSerGluAlaThrGlyAla---ProThr----- 807
Db 35960 ACCGCGCGGTCTTTGCGCGGGGCGCTTGGCTTACACGCTCGGCGCTGCAAGGCCCGCG 36019
Qy 808 -----ProProAlaProProSerProSerAlaProProValValProLy 824
Db 36020 CTCTCGGTGACACCGCT 36073
Qy 824 sGluGluLysGluGluThrAlaAlaAlaProProProValGluGluGluGlnLy 844
Db 36074 CAAGCCCTCGCAGGAGCGAGTGCAACCTCGCCCTCGCGCGCGCGCTCTCGCTCATGGCC 36133
Qy 844 sProProAlaAlaGluGlu-----LeuAlaValAs 854
Db 36134 TCCCCCGGGCTCTTCGTGCTCTTTTCCCGCATGCGTCTTTGGCGCCCGATGGCGCGTCC 36193

```
FT /note= "gene product is highly homologous to type I PKs
FT that are known to be involved in the synthesis of
FT polyketide compounds"
FT 942..7115
FT /tag= c
FT /note= "module 1 of SorA"
FT 7203..12884
FT /tag= d
FT /note= "module 2 of SorA"
FT 13455..13616
FT /tag= e
FT /note= "module 3 of SorA"
FT 19870..24556
FT /tag= h
FT /note= "module 1 of SorB"
FT 19871..46318
FT /tag= f
FT 19871..46318
FT /tag= g
FT /product= "SorB"
FT /note= "gene product is highly homologous to type I PKs
FT genes"
FT 24638..30820
FT /tag= i
FT /note= "module 2 of SorB"
FT 30881..35446
FT /tag= j
FT /note= "module 3 of SorB"
FT 35528..40114
FT /tag= k
FT /note= "module 4 of SorB"
FT 40190..46318
FT /tag= l
FT /note= "module 5 of SorB"
FT 46851..47891
FT /tag= m
FT /product= "SorM"
FT /note= "gene product is homologous to the
FT methyltransferase from Streptomyces hygroscopicus that is
FT involved in the synthesis of the polyketide rapamycin"
FT
FT
XX US5716849-A.
XX
XX 10-FEB-1998.
XX
XX 14-DEC-1996; 96US-00764233.
XX
XX 24-AUG-1993; 93WO-US007954.
XX 08-JUN-1994; 94US-00258261.
XX 09-OCT-1996; 96US-00729214.
XX
XX (NOVS ) NOVARTIS FINANCE CORP.
XX
XX Ligon JM, Hill DS, Neff S, Beck JU, Ryals JA, Schupp T;
XX
XX WPI; 1998-158369/14.
XX
XX DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
XX biosynthesis of soraphen useful as antimicrobial agent against
XX phytopathogenic fungi.
XX
XX Claim 2; Col 47-90; 64pp; English.
XX
XX The present sequence contains a cluster of genes that encode polyketide
XX synthases (PKSs) that are involved in the synthesis of soraphens in
XX Sorangium cellulosum. The proteins encoded by the present sequence are
XX SorA, SorB, SorC and SorM. SorA and SorB contain biosynthetic modules
XX which contain a beta-ketoacylsynthase, an acyltransferase, a
XX ketoreductase and an acyl carrier protein domain, as well as a beta-ketone
XX processing domains. S. cellulosum soraphens are useful as a cytostatic
XX and antimicrobial agent active against phytopathogenic fungi. Soraphen-
XX producing transgenic plants or biological control agents can also be
XX produced, which may reduce crop losses and nutritional deprivation for
```

```
CC local populations in many parts of the world. (Updated on 17-OCT-2003 to
CC standardise OS field)
XX Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 U; 0 Other;
XX
Alignment Scores: 1.16e-10 Length: 49377
Pred. No.: 641.00 Matches: 661
Score: 31.16% Conservative: 298
Percent Similarity: 21.47% Mismatches: 1183
Best Local Similarity: 4.85% Indels: 941
Query Match: 2 Gaps: 140
DB:
US-09-522-753-5 (1-2517) x AAV05287 (1-49377)
QY 4 SerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrProHisSer 23
Db 33591 TCACGCGCTCAGCGCCAGCTCCACGCGACGCGACGAGCGCGCCCTCGCCC 33650
QY 24 LeuSerTyrProValGlnIleAlaArgThrHisThr----- 35
Db 33651 TGTCTCTCCACCTCTCGAGCTTTCACCACGCGCCAGAGACGACGCGTGACA 33710
QY 36 -----AspValGlyLeuLeu-GluTyrGlnHisHisSerArgAspTyrAl 50
Db 33711 CCTGGCGCTACCGCATCAGTGAGGCTCTGACACCGCGCCGACGCGCGACCTCG 33770
QY 50 aSerHisLeuSerProGlySerIleIleGlnProGlnArgArg----- 65
Db 33771 CCGGCACCTGGCTCTCGTCCGCGTCCGCGTCCGCGACGAGCGCTCTCTGCCACGC 33830
QY 66 -----ProSerLeuLeuSerGluPheGlnPr 74
Db 33831 TCACCGATGCGCTTACCGCGCGCGCGGTCTCTCGCGCTCGCGCTGAG-----CC 33884
QY 74 oGlyAsnGluArgSerGlnGluLeuHisLeuArg----- 85
Db 33885 AGGTTACATAGCGCGCGGCTCTCACCAGCAGCCTCGCGAGGCTTGTTCGCGAGACTG 33944
QY 86 -ProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGlu 105
Db 33945 CCGCGATTCCGCGGTCTCTCCCT----- 33969
QY 105 rLysArgProArgLeu-----GluLeuLeuProAspPr 116
Db 33970 -CCTGGCCCTCGACGAGCGCCCTCGCGACCATGCGCCCTCGCGCGGCTGCCC 34028
QY 116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer---GluAspLe 135
Db 34029 TCTGCTCGCCCTCGTCCAGCCCTCGGCGACCTCGGCCCTCGAGGCTCCCTTGTGGCTCT 34088
QY 135 uThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHi 155
Db 34089 TCACGCGCGCGCGCTCTCGATTGGACACTCCGACCATCTGCCCATCCACCCAGGCCA 34148
QY 155 sThrAspProGluLeuGluLeuValProArgLeuSerLysGluLeuIleGlnAs 175
Db 34149 T-----GATCTGGGCTTGGCGCGCTCGTCCGCTCGA----- 34182
QY 175 nMetAspArgValAspArgGluIleThrMetValGluGlnIleSerLysLeuLysLy 195
Db 34183 -----GCACCCGAGCGGTGGCGGTCTGCTCG 34211
QY 195 sLysGlnGlnLeuGluGluAlaAlaLysProGluProGluLysProValSe 215
Db 34212 ACCTCGCGGAGCGCTCGACGCGAGCGCGCGCGCTGCTCCCGCGCTCCCGCAGC 34271
QY 215 rProProLleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235
Db 34272 GCCA-----CGACCG 34280
QY 235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255
Db
```

| | | | |
|----|-------|---|-------|
| Qy | 2014 | ----- | 2014 |
| Db | 24034 | GACCCGAGCGCGAGCCCAAGAGCTACGTCGCGCATGCGCGATGCTCGACCGAGATCGAC | 24093 |
| Qy | 2015 | -----ProLaSerAlaSerAapProHisArgGluIuYThrGlnSerLysPr | 2030 |
| Db | 24094 | CTCTTCGACCCCTGCTTCTTCGCGCATCAGCCCGGAGGCCAAACACC---TCGACCCC | 24150 |
| Qy | 2030 | oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTYrHisGlySerSerTYrSerPr | 2050 |
| Db | 24151 | CAGACCGGCTGCTCTCGAATCTGCTGGCTGGCCCTCGAGGACCGCGGCATCGTCCCC | 24210 |
| Qy | 2050 | oGluGlyValGluProValSerProValSerSer----- | 2061 |
| Db | 24211 | ACCTCCCTCAAGACTCCTCTCAGCGCGCTTCCTCGTCGGCATCTGCGCGCGGAATACGCG | 24270 |
| Qy | 2062 | -----ProSerLeuThrHisAepLysGlyLeuProLysHi | 2073 |
| Db | 24271 | ATGCAAGAGCGAGCTCGGAAGGTTCCGAGGTTTACTTCATCAAGGACATCCCG----- | 24325 |
| Qy | 2073 | sLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly-- | 2092 |
| Db | 24326 | -----CGTCCTTTGGCGCGGGGGCG | 24345 |
| Qy | 2093 | -----ProValLysLeuGlyGlyAlaAlaHisLeuProHisLeuLeuArgProLeuProGl | 2111 |
| Db | 24346 | TTGGCCTATACGCTCGGGCTCAGGGGGCGC-----GATCTTCGGTCGACACCGCC | 24396 |
| Qy | 2111 | uSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr | 2131 |
| Db | 24397 | TGCTCTCTCTCGCTCTCTCCC-----TCCACCTCGCTCGCAAGCCTCCGACAGGCG | 24450 |
| Qy | 2131 | gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTYrThrArgHisHi | 2151 |
| Db | 24451 | GATGCAACC----- | 24460 |
| Qy | 2151 | sProGlnGlnLeuSerAlaProLeuProAlaProLeuTYrSerPhe-----Pr | 2167 |
| Db | 24461 | -----TCGCCCTCGCGGGCGGTGCTCGCTCATGGTCTCCCCCAGACC | 24504 |
| Qy | 2167 | oGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTYrLeuProPr | 2187 |
| Db | 24505 | TTCGTCATCTTTCCCGTCTGCGCGCTTGGCGCGCCGACGGCC-----GCTCCAAG | 24555 |
| Qy | 2187 | oProAapHisGlyAlaProAla-----ArgGlySerProHisSerGl | 2201 |
| Db | 24556 | ACCTTCTCGGACAAACCCACCGCTACGACGCGGAGAGGCGTCTGCTCTGCCCTC | 24615 |
| Qy | 2201 | uGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGl | 2221 |
| Db | 24616 | GACGGATCGGACCGCTCTCGCCGGAGACACCGCG-----TCCTC | 24657 |
| Qy | 2221 | yIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVa | 2241 |
| Db | 24658 | GTCTCTGTCGCGGCACCGCATCAACACGACG-----GCGGTGAGCGGT | 24705 |
| Qy | 2241 | lTYrProLeuLeuTYrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe | 2261 |
| Db | 24706 | ATCACCG-----CCCCCAACGGCACCTCCACACAGA-----AGTC | 24741 |
| Qy | 2261 | rProGlyAanThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAanSerAl | 2281 |
| Db | 24742 | CTCGGGCGCGCTCCACAGCGCCGCATCACCCCGCGCGAGCTCGAGTGTGTGAGTGC | 24801 |
| Qy | 2281 | aMetValLysSerLysGlnGluIleAsn----- | 2291 |
| Db | 24802 | CATGGCACCGGCACCTCGCTGGGAGACCCCATCGAGGTGCAAGCCTGGCGCGTCTAC | 24861 |
| Qy | 2292 | -----LysLysLeuAanThrHisAanArgAanGluProGluTYrAs | 2305 |
| Db | 24862 | GCGGAGGCGAGCCCGCTGAAAGACCTCTCTTCTCGGCGCGCTCAAGACCAATCGGC | 24921 |

| | | | |
|-----------|---|--|-------|
| Qy | 2305 | nIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMe | 2325 |
| Db | 24922 | CATCTCGAGCGG-----CCTCGGCGCTC-GCGGGCGTCCGCAAGT | 24952 |
| Qy | 2325 | tThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaI | 2345 |
| Db | 24963 | GGTC-----GCCTCGCTCGCCACGACGCCCTGCCCCACCCTCCACGCGAC | 25010 |
| Qy | 2345 | eIleArgLysAlaLeuMetGlyLysTyAspGlnTrpGlu-GluSerProProLeuSerA | 2365 |
| Db | 25011 | CCACGCGATCCCTCATC-----GAGTGGGAGGCGCTCGCATCGAGCTCGT | 25058 |
| Qy | 2365 | laAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrA | 2385 |
| Db | 25059 | CG-----ATACCCCGAGCGCTTGGCCCGCCACGACGATGCGAGTCCGCCCGCGCGC | 25109 |
| Qy | 2385 | laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLys----- | 2402 |
| Db | 25110 | CGGCArTCTCGGCTTCGGATCTTCGGGCAACCAACGCCACGTCATCTCGAAGAGGCTCC | 25169 |
| Qy | 2403 | -----AlaLysValSerGlyAArgProSerSerArgLysAlaLysSer---ProAlaP | 2419 |
| Db | 25170 | CGCGCGCTTCGGCGCGAGCCGCCACCTCACAGCGGCGTCCGACGCGCTCCGCGCGC | 25229 |
| Qy | 2419 | roGlyLeuAlaSer-----GlyAspArgProProSerValSerSerValHisSerG | 2436 |
| Db | 25230 | GTGGCCGCTGCTCTGTCTGCGGCAGGAGCGAGGCGCGC-----TCCGCGCCAGCGAA | 25283 |
| Qy | 2436 | luGlyAspCysAsnArgArgThrProLeuThr-----AsnArgValTrpGluAspA | 2453 |
| Db | 25284 | GCGGCTCCCGACCACTCGTCGCCCGACGACACCTCACCTCGCGGATGTGGCTATTC | 25343 |
| Qy | 2453 | rgProSerSerAlaGlySerThrProPheProTyAsnProLeuIleMetArgLeuGlnA | 2473 |
| Db | 25344 | GC----- | 25345 |
| Qy | 2473 | laGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaG | 2493 |
| Db | 25346 | -----TGGCCACCACCGCGCCCTTCGAGCACCAGCGCGCTCTCGT | 25388 |
| Qy | 2493 | lyProHisHisAlaTrpAspGluGluProLys | 2503 |
| Db | 25389 | AGCCCAACACCGCGAGCTCTCTCTCGCGC | 25420 |
| RESULT 56 | | | |
| Id | AAV05287 | | |
| XX | AAV05287 standard; DNA; 49377 BP. | | |
| XX | AAV05287; | | |
| DT | 17-OCT-2003 (revised) | | |
| DT | 21-MAY-1998 (first entry) | | |
| XX | The soraphen biosynthesis gene cluster from Sorangium cellulosum. | | |
| XX | Polyketide synthase; PKS; biosynthesis; soraphen; SorA; SorB; SorM; | | |
| KW | biosynthetic module; beta-ketoacyl synthase; acyltransferase; | | |
| KW | ketoreductase; beta-ketone processing domain; cytosolic agent; | | |
| KW | antimicrobial agent; phytopathogenic fungi; transgenic plant; | | |
| XX | biological control; ss. | | |
| XX | Polyangium cellulosum. | | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | CDS | 383..760 | |
| FT | | /tag= a | |
| FT | | /product= "SorA" | |
| FT | | /note= "gene product highly homologous to the reductase | |
| FT | | domains of type I PKSs such as eryA from | |
| FT | | Sacccharopolyspora erythraea" | |
| FT | CDS | 927..19874 | |
| FT | | /tag= b | |
| FT | | /product= "SorA" | |

| | | | | | | | |
|----|-------|--|-------|----|-------|--|-------|
| Db | 21921 | GCGETTCTGGAGGCCATCGAGACGGGAATATCGACGCGCTCAGCGGCCAGCTCCACGT | 21980 | Db | 22964 | TCGACGCCGAAGGGTCGACAGTCCGCGCGTGTTCACACGGGCGGCATCGGCGCCACG | 23023 |
| Qy | 1483 | -----ThrProProValHisProLeuA | 1491 | Qy | 1776 | -----ProLeuSerProGlyGlyProThr- | 1783 |
| Db | 21981 | GGACGGCGACGACGCGCGCCCTTGCCCTCTCTCCACCCCTCGCGAGCTTCG | 22040 | Db | 23024 | CTCCGCTCGCGCCACTCTCTCATGGAGCTCGCGACGCTTGTCTGTGCAAGGCTCTAG | 23083 |
| Qy | 1491 | spValMetAlaAspAlaArgAlaLeu-GluArgAlaCysTyrGluGluSerLeuLeuSer | 1510 | Qy | 1784 | -----HisLeuThrLysProThrThrThrSerSerSerGlu | 1795 |
| Db | 22041 | CCACGAGCGGCAAGACGAGGCGACGCTCGACGCGCTGCGCTACGCGCATCGTGGAGCC | 22100 | Db | 23084 | GGCAGGGAACCTCCACGACCTCGTCTCGGTCCTCGACCCCTCGAGCGCTTCTCTTTCT | 23143 |
| Qy | 1511 | -----ArgProGly-----ThrAlaSerSerSerGly | 1519 | Qy | 1796 | Arg-GluArgAspArgAspArgGluArgAspArgGluArgGluSerSerLeu | 1815 |
| Db | 22101 | TCTGACCAACCCACACGCGCGCGCTCGCGGCGGACCTGCGTCTCTGTCGCGCG | 22160 | Db | 23144 | CGTCCATCGAGCGCTCGGCGCGGACAAACAGCGGATACGCGCGGGAACGCT | 23203 |
| Qy | 1537 | GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro-----Arg | 1537 | Qy | 1815 | uThrSerThrThrValGluHisAlaProIleTrp | 1827 |
| Db | 22161 | CGCTCTGAGACGACGCGCTCCCTCGCGCTCACCGAGCGCTCGCGCGCGCGC | 22220 | Db | 23204 | TCCTGACGCCCTGGCGGACCGAGCGCGAGTCTTGAGACAGCGCGGACGTCGCTGGTGT | 23263 |
| Qy | 1538 | GlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly | 1557 | Qy | 1828 | -----ArgProGlyThrGluGlnSerSe | 1835 |
| Db | 22221 | CGCGGCTCTCGCGCTGCGCTGAGCCAGCCACCTCGA-----CGCGAGGC | 22268 | Db | 23264 | GGGCGGTGGGCGGCGGCTGATATTACAGGGGCGCTTGCAGCGCCAGCTGGAGC | 23323 |
| Qy | 1558 | SerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSer | 1577 | Qy | 1835 | rgLysSerGlySerSerGly-----GlyGlyGlyGlySerSerSerArgProAl | 1852 |
| Db | 22269 | TCT-----CGCGAGACCTCGCGGAGCTTGGCCGAGACCGCGCGCTCGCGG | 22319 | Db | 23324 | AACGTGCTCTGTCGCGATGGCCCTTCGTGGCGCTGCGCGCGCTCGCGCAAGCCCTGG | 23383 |
| Qy | 1578 | LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis | 1597 | Qy | 1852 | asr-----HisSerHisAlaHisGlnHis | 1860 |
| Db | 22320 | CGTCTCTCGCTCTCGCTCGCGGCGGCGGCTCGCGGCGGCGGCGGCGGCGGCGG | 22379 | Db | 23384 | AGCAGCAGACGACCGCTCACCGTCGCGACATCGACTGGGCGCGCTTTCGCGCTCGA | 23443 |
| Qy | 1598 | SerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGly | 1617 | Qy | 1861 | -----SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValle | 1878 |
| Db | 22380 | GGGACTCGCTCTCGCTCACCTCGTCCA---AGCCCTCGCGGACATCGCCCTCGAGC | 22436 | Db | 23444 | TCAGCGTCTCGCTCGCGCGCTCTCGCGAGCTTTCGCGAGCAGCGC | 23492 |
| Qy | 1618 | ValSerGlyValAspLeuThr-----ArgSerHisIle | 1628 | Qy | 1878 | uHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValle | 1898 |
| Db | 22437 | GCCTT---GTGGCTCTCACCGCGCGCGCTCTCGTTCGACACTCGACCCCATCGC | 22493 | Db | 23493 | -----GCCCTCGAAGAC----- | 23504 |
| Qy | 1629 | Pro-----LeuAlaPheAspProThrSerIleProArgGlyIlePro | 1642 | Qy | 1898 | uArgSerThrSerThrSerProValArgProAlaAlaThrPheProProAlaThrHi | 1918 |
| Db | 22494 | CCATCCGACGAGCGGATGAGCTGGGCGCTGGCGCGCTCGTTCGCGCTCGAGCA---CCC | 22550 | Db | 23505 | -AGAGAAGGCGCTCTCTCTCCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 23563 |
| Qy | 1643 | LeuAspAlaAlaAlaTyrTyrLeuProArgHis | 1654 | Qy | 1918 | sCysProLeuGlyGlyThrLeu----- | 1927 |
| Db | 22551 | CGAGCGTGGGAGGGCTCGTTCGAGCTCGCGCGAGCATCGACGGAGCGCGCTGGGCGG | 22610 | Db | 23564 | GGAGCGCTCGAGAGCGAGCAGCTCGTCTCGCGCGCTGGTGTGCGACGAGCGG | 23623 |
| Qy | 1655 | LeuAlaProAsnProThr-----TyrProHisLeuTyrProPro-----Tyr | 1668 | Qy | 1927 | yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPro---ArgValAl | 1946 |
| Db | 22611 | CTTGCTCCGGTCTCGCTCGCGCTGGCAACGATGAGGACGAGCTCTCTCGCGCGCGCGG | 22670 | Db | 23624 | CCCTCGCTCGCGCCAGAGGCGG-----CTTCCAGCTCGACCCCGACAGGCTTCT | 23677 |
| Qy | 1669 | LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleLeuAsnAsp | 1688 | Qy | 1946 | aArgProGluArgProArgAlaAsp-----ThrGlyHisAlaPheLeuAlaLys | 1962 |
| Db | 22671 | GTTCTACGCTCGCGCTCTCGCGCGCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG | 22724 | Db | 23678 | TCGACCT---CGTCTCGATTCGATCATGACCGTCGAGCTTCGTGCGCGCTTGCAAC-AG | 23733 |
| Qy | 1689 | TyrIleThrSerGlnGln-----MethHisAsnThrAlaThrAlaMetAlaGlnArg | 1706 | Qy | 1962 | sProProAlaArgSerGlyLeuGluProAlaSerSer----- | 1974 |
| Db | 22725 | TACCTTAAGCCCGAGGACCTCTCTATACCGAG-GACCGCGCGCGCTGGCGCTC | 22783 | Db | 23734 | GCCACCGCATCAAGCTCCCGCGCACCTCGCCCTTCGACCATCCCTCTCTCTCATCGCGTC | 23793 |
| Qy | 1707 | AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAla | 1726 | Qy | 1975 | -----ProSer-----LysGlySerGluProArg-- | 1982 |
| Db | 22784 | AGTCGCGCGATGGCTCGCTCGAAGGCGGAGCAGCAGCTCGCTCATCAGCGCGCGG | 22843 | Db | 23794 | CGGCTCTTCATGCGGAGCTCGTCCGCCACCGCTCGGACGAGGCTCTCCGCGAGCGG | 23853 |
| Qy | 1727 | GlyProArgGlyIleLeuAspLeuSerGlnValProHisLeuProValLeuValProPro | 1746 | Qy | 1983 | -----ProLeuValProValSerGlyHisAlaThrIleAlaArgThrPr | 1998 |
| Db | 22844 | GGGCGGCGGCGGCGCGCTCGAGGCTCCACGCGAGCTCACGCGCTCGCGCGCGCGCGG | 22903 | Db | 23854 | ACGCGCGCGCTCCGCGCGCTCGAGCGACGAGCGCATCGCATCGTGGCATGGCC | 23913 |
| Qy | 1747 | ThrProGly-----ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPro | 1762 | Qy | 1998 | o----- | 1998 |
| Db | 22904 | TCACCTTCGCGCGTGTGATGTCGCGGACAGGAGCGCTCTCGACGCTTCTCGAGCAGC | 22963 | Db | 23914 | CTGCGCTCGCGCGCGCGCTCGCGGATGTGACGCTCTTTGGGAGTCTCTCCCAAGGG | 23973 |
| Qy | 1763 | ThrAlaProGln-----ProPheSerArgHisSerSer----- | 1775 | Qy | 1999 | -AlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaPro----- | 2014 |
| | | --- | | Db | 23974 | CGCAGCGGTCGAGCCCATTCACAGAGCGCGCTGGGAGCGCGCTGCGCTCTACGACCC | 24033 |

874 oAlaLysGlyLysAspAlaGluAlaAlaGluAlaThraAlaGluGlyAlaLeuLysAlaG1 894
19951 CTCAGCGGCTCGCGAGCCCTCGCCGAGGACACCGCGTCTCGCCCTCGTCCGCGC 20010
894 uLysLysGluGlyGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGlnAs 914
20011 ACCGCCATGAACCATGACGCGGCTCGAGCGCATCACCCGCCCAATAGCAGCTCCAC 20070
914 pSerAspSerSerAlaThrCysSerAlaAspGluAlaValAspGluAlaGluGlyAspLy 934
20071 CAGAGGTCTCTCGCGCGCGCTCCACGACGCCCAATATCGCCCTGCCGAGCTCGAGTC 20130
934 sAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAs 954
20131 GTCGAATGCCATGGCAGCGGACCTCTCTGGAGAGCCCATCG-----AGGTCAA 20181
954 nAla----SerProGlnLysProLeu-----AspLeuLysGlnLeuLysGlnArgAlaAl 971
20182 GCCCTGGCCCGCTCTACGCGCATGGCAGACCCGCTGAAAGGCTCTCTCTTCGCGCA 20241
971 aAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaPr 991
20242 CTCAGACCAACATGTGCCATCTCGAGG-----CGCCCTCCGCGCTCGCGGCGTCGCC 20295
991 oThrLysProAlaPro-----ProAlaProProPro-----ProGlnAs 1004
20296 AAGATCGTCGCTCCCTCCGCGCATGAGCGCTCGCCGCCACCTCCACACAGCCCGCGC 20355
1004 nLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer-----ProArgGlyLy 1022
20356 AATC---CCCTGATCGAGTGGGATGCGCTCGCCATCGACGCTCGATGTCACGAGGCG 20412
1022 sSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe 1042
20413 TGGG-----CCGCGCACGAAGTGGCAGTCCCGCGCGCGCGCGTCTCC 20457
1042 uProGlyAspProProCysTrpThrSerGlyLeuProPhe----- 1055
20458 GCCTTCGACTCTCCGCGACCAACGCGCCACGTTATCTTCGAAGAGGCTCCCGCATCCCG 20517
1056 ----ProValProArgGluValIleLysAlaSerProHisAlaProAspProSerAl 1074
20518 CAGCGAGGCGCCACCGCGGACAGCTCGCTCGCAGC---CGCTTCGCGAGCTGGCCC 20574
1074 aPheSerTyrAlaPro-----ProGlyHisProLeuProLeuGlyLeuHisAs 1090
20575 GTGCTCTCTGCGGAGGAGCGCGCGCTGCGCGCCAGGCGCCAGAGCTCCGCGAC 20634
1090 pThrAlaArgProValLeuProArgPro-----ProThrIleSerAsnProPr 1106
20635 CACTCTCTCGCCACGACGACCTCGCCCTGGCGATGTAGCTACTCGCTCGCCACCAACC 20694
1106 o-----ProLeuIleSerSerAlaLysHisProSerValLeuG1 1119
20695 CGGGCTACCTTCGAGCACCGTGGCGTCTCGTGTGCCACCGCGGAAGAGCTCTCTCC 20754
1119 uArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerG1 1139
20755 GCGTCTGATTGCTCGCCAGGAGGAGCCGCGCC-----CGAGCACCTGCTCGCAA 20805
1139 uHisAlaLysAlaPro-----ValGlyPro-ValThrMetGlyL 1152
20806 CGAAGCGGAAGCCACGCAAGGTCTGCTCTTTCTTCTGGGCAAGGCTCGCAGTGGAA 20865
1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGly----ValLysGlnGluG 1171
20866 GGGATGGCCCTCTCCCTGCTGATACCTCGCGGTCTTCGCGGACAGCTCGAAGCGTGC 20925
1171 lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG 1191
20926 GAGGCGCCCTTCGCGCGCCACGAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 20978

1191 luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG 1211
20979 -----CGAGGA-----GGCGCGCCCGCTCGACCGGTCGAGCTG 21015
1211 lYlIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrH 1231
21016 GT-----CGAGCCCGCTGTCTCGATGATGCTCTCGCTGGCGCCCTGTGCGCGTCC 21069
1231 isGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspS 1251
21070 ATGGCGTCTGAGCC-----CGACGGGTGGTGGCCAT 21102
1251 exProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrG 1271
21103 AGCAGGGGAGATCCCGCGC-----CTGTGTGGCGCGCGCTGTGCTC 21150
1271 luGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerL 1291
21151 GAGGACGC-----TGCCAAGCTGGTGGCGTGGCAGCCGTGGCTC 21192
1291 yS-----GluAspGlyArgSerSerS 1298
21193 GTGAGCTCGCCGCGCGGCGCATGGCGCGGTGGAGTGGCGGAGCGCGAGGTCGCA 21252
1298 erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV 1318
21253 CGCGCGCTCCA-----GCGCTATGGCATGCGGCTC 21282
1318 aIcGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProG 1338
21283 TCATCGGGGCGATCAACAGCCCTCTTTACAGCAGCATCTCCGCGGAGCCCTCGCGTC 21342
1338 luArgHisSerProHis-----HisLeuLysGluGlnHisIleArg- 1352
21343 GCGCGCTCTCGCGCATCTGGAGTCCGAGGCGTCTTCGCCCTCAAGCTGAGTACGAC 21402
1353 -----GlySerIle---ThrGlnGlyIleProArgSerTyrValG 1365
21403 TTGCGCTCCCACTCCGCGCAGGTGAGTCGATTTCGCGAGAGCTCTCGATCTCTGTCG 21462
1365 luAlaGln-----GluAspTyrLeuArgA 1373
21463 TGCTCGAGCGCGCTCGACGCGGTCCCGTCTTACTCCAGGTGAGCGCGCGCGATC 21522
1373 rgGluAlaLysLeuLysArgGluGlyThrProProProPro----- 1387
21523 GACGGGAGCGAGTCCAGCGCGCTACTGTGTACCGGACCTCCGCGAGCGGTCCGCTTC 21582
1388 -----ProProSerArg----- 1391
21583 GCAGAGCTGTGCAAGGCTCTTTCGCGGAGAACATCGCTTCTCGTGGAGGTGAGCCCC 21642
1392 -----AspLeuThrGluAlaTyrIleThrGlnAlaLeuGlyProLeu- 1405
21643 AGTCTGTGCTGACCTTGCGCTTTCGAGCTCTCGAGCGTCCGAGCGCTCGCGCGCG 21702
1406 -----LysLeuLys-----ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
21703 GTGGTCTGCTCTGTGGAGCGAGGAGGATCTACGCGCTTCTCGTCTGCTCTCC 21762
1423 rgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaP 1443
21763 GAGCTCTAGTCAACGCGCTTCGCGCTTGGATTGGACGACGATCCT-----GCC 21810
1443 roArg-----ProLeuLysGluGlySerIleThrGlnGlyT 1455
21811 CCGGGAAGCGGTGCGCTGCGCCACTACCCCTTCC-ACGCGAGCGCTTCTGGCTCGA 21869
1455 hrProLeuLysTyrAspThrThrGlyAlaSerThrThrGlySerLysLysHisAspValArgS 1475
21870 CGCTTCCAGCGACCGCGCGCGCTCAACCACTTGTCTCCGCTCG-----AGGG 21920
1475 erLeuIleGlySerProGlyArg----- 1482

| | | | | | | |
|----|-------|---|----|-------|--|-------|
| Db | 18020 | GCCTCGTCCGCGCCCGCTCGG----- | Db | 18887 | -----ACAAAGCTCCGAAACCGCTCGGAGAGCGAGCAGATCCACCTG | 18927 |
| Qy | 275 | aMetArgLysIleuLeuTyrPheLysArgAtgAsnHisAlaArgLysGlnTrpLys | Qy | 584 | -----ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIle-----ThrProGlnG | 600 |
| Db | 18050 | CTGCGCTCGCGGCTT-----CATGCCCGAGGACCATCTCTCATCCCGTGGTA | Db | 18928 | CTCTCTCGTCTGTCGCGCACGAAGCGCCCTCTGCTGGGCCATACCGACGCTCCCGAG | 18987 |
| Qy | 295 | sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuLys-----Ly | Qy | 600 | nSerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTrpThrGluGluGluMe | 620 |
| Db | 18101 | CCGCGCCCATTTGGCGCTCAGCTGCCGCGATGGCTCGTCAAAAAGGCGCTGAGCACCTCG | Db | 18988 | GTGACACCC-----ACAAGGGCTTCATGACCTCGGCCTCGAT | 19026 |
| Qy | 312 | sValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgLuty | Qy | 620 | tGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMe | 640 |
| Db | 18161 | TCCTCATCAGCCGACGAGGGCGCCAGGCGCGAGCGCTGGAGCTCCACGCGCGACTCA | Db | 19027 | TGCTCATACCGCTCGAGCTTCGTC-----GGCGCTTCGACGAGGCC | 19068 |
| Qy | 332 | rTyrGluLysGlnPheProGluIleArgLysGlnArg-----GluLeuGlnGluArgMetG | Qy | 640 | tValGlySerLysThrValSerGlnCysLysAsnPheTyr-PheAsnTyrLysLysArgG | 660 |
| Db | 18221 | CGCGCTCGCGCGCGGTACCTTCGCGCGTCCGATGTCGCCGACAGGCGCTGTCTG | Db | 19069 | ACCGGCATCAAGCTCCCGGCCACCTCGCTTCGACCATCTCTCTCTCATCGCTCGCG | 19128 |
| Qy | 351 | n-----SerArgValGlyGlnArg-GlySerGlyLeuSerMetSer----- | Qy | 660 | lnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaA | 680 |
| Db | 18281 | CCACGCTTCTCAGCAGCTCAGCCCGGA-----GGGCCACAGGTGAGCGCGCTGTCCACG | Db | 19129 | CTCTCTTTCGCGA---CTCGCTCGCCACGCCCTCGCGCGAGGCTCTCCGTCGAGCGC | 19185 |
| Qy | 365 | -----AlaAlaArgSerGluHisGluValSerGluI | Qy | 680 | rArgLysLysLys-LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProVal | 699 |
| Db | 18338 | CGGCGGCATCGAGCCCCACCGCTCGCTCGCGCCACCTTCATGAGGATCTGCCGAGG | Db | 19186 | GAGCGCGCGCTCCCGCGCTTCGCTCGCGAGCGAG-----CCCATCGCC | 19236 |
| Qy | 375 | leIleAspGlyLeuSerGluGlnGluAsnLeuLysGlnMetArgGlnLeuAlaValI | Qy | 700 | valGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGlu | 719 |
| Db | 18398 | TTGCTCCGCGCAAGGTACAAGT-----GCAAGACACCTCCACGACCTGCTCGCT | Db | 19237 | ATCGTCGGCATGCGCTTCGCTTCGCGCGGCGCATCGCGATCTCGACGCTCTTTGGAG | 19296 |
| Qy | 395 | leProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyL | Qy | 720 | -----GluAlaGluAlaLeu---HisAla-SerGlyAsnGluVa | 731 |
| Db | 18449 | CTCGACCCCTC----- | Db | 19297 | TTCTCGCCCAAGACGCGAGCGCTCGAGCCCATTCCTCCCATGCCGATGGGATGCGGT | 19356 |
| Qy | 415 | eumetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluG | Qy | 731 | lProArg-----GlyGluCysSerGlyProAlaThrVa | 742 |
| Db | 18460 | -----GACGCTTTGTTCTTCTCTCTCGTCCGCGCGT-----GTCTGGGCGCGC | Db | 19357 | GCCTCTACGACCCGACCCGACGCGCAAGCGCCAGAGTACCTCCGCGATGCGCGCATG | 19416 |
| Qy | 435 | lnGluLys-----GluThrPheArgGluLysPheMetGlnHisProL | Qy | 742 | lAsnAsnSerSerAspThrGluSerIleProSer-----ProHi | 755 |
| Db | 18506 | GACAAACAGGCGGTATCGCGTGGGAACGCTTCTCGATGCCCTGGCCGAGCAGCGCG | Db | 19417 | CTGACCGAGTCACTCTTCGATCTCTCTCTTTTGGCATCAGCCTCGCGAGGCCAA | 19476 |
| Qy | 449 | ysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeut | Qy | 755 | sThrGluAlaAlaLysAspThr---GlyGlnAsnGlyProLysProProAlaThrLeuG | 774 |
| Db | 18566 | GCAGCTTGGGCTGACGCGACA-----TCGGTGGCT | Db | 19477 | TACTCGACCCCGACGACCGCTCTCTCGAATCTGCTGGCTGGCTCGAGGACGCC | 19536 |
| Qy | 469 | yTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer----- | Qy | 774 | yAlaAspGlyProProGly-----ProProThr----- | 784 |
| Db | 18599 | GGGG-CGTGTGGGCGCGCGCGCATGGCTACCGGCTCTTGGCAGCCAGCTAGAGCAA | Db | 19537 | GGCATCTCCCTCCACCTCAAGGATCTCCACCGGCTCTTCGTCGCGCATCGCGCC | 19596 |
| Qy | 482 | -----LeuValArgArgSerTyrArgArgArgLysSerG | Qy | 785 | -----ProProArgArgThrSerArgAlaProIleGluPr | 796 |
| Db | 18658 | CGCGTCTGTCGCGATGCCCTCGTGGCGTGGCGACGCTCGCTGGCGCTGGCGCTGAG | Db | 19597 | AGCGAATACGACTGCGAAACACGAGCTCGAAAGAGTCAAGGCTATGCGCTCCAAGGC | 19656 |
| Qy | 494 | lnGlnGlnGlnGlnGlnGlnGln-----G | Qy | 796 | oThrProAlaSerGluAlaThrGlyAla---ProThr----- | 807 |
| Db | 18718 | CACGAGGACACCTCACCCTGCGCGCATCGACTGGCGCGCTTTGGCTTCGTC | Db | 19657 | ACCGCGGCTCTTTGCGCGGCGCTTGGCTTACACGCTCGCGCTGCAAGGCCCGCG | 19716 |
| Qy | 505 | lnGlnGlnGlnGlnGlnProMetPro-ArgSerSerGlnGluLysAspGluLysGlu | Qy | 808 | -----ProProAlaProProSerProSerAlaProProProValProLy | 824 |
| Db | 18778 | AGCGCGCTCGCTCCCGCGCTCTCTGCGGATTTGCCGAGCGCGCGCTCTCGAA | Db | 19717 | CTCTCGGTGACACCGCT | 19770 |
| Qy | 525 | LysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAs | Qy | 824 | sGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGluGluGlu | 844 |
| Db | 18838 | GCCAGCGCGATCGCTCCAGCAAGACGGGCGCACAGGCTCTCTCTG | Db | 19771 | CAAGCCTCCGACAGGCGGAGTGCAACCTCGCCCTCGCCCGGCGGCTCTCGTCATGGCC | 19830 |
| Qy | 544 | pLeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaVa | Qy | 844 | sProProAlaAlaGluGlu-----LeuAlaValAs | 854 |
| Db | 18886 | ----- | Db | 19831 | TCCCCCGGCT | 19890 |
| Qy | 564 | lAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThr--- | Qy | 854 | pThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyPr | 874 |
| Db | 18891 | AAGACCTTCTCGACCAACGCGCGGTACTCGGACGCGGAGGCGCTCGCTCTCTCTG | Db | 19891 | AAGACCTTCTCGACCAACGCGCGGTACTCGGACGCGGAGGCGCTCGCTCTCTCTG | 19950 |

Qy 1878 uHisAsnThrGlyMetLysGlyIleThrAlaValGluProSerLysProThrValLe 1898
Db 23493 -----GCCTCGAGAC----- 23504
Qy 1898 uArgSerThrSerThrSerProValArgProAlaAlaThrPheProProAlaThrHi 1918
Db 23505 -AGAGAGGGCGCTCTCTCCGAGACGGCGCGCCCGCCCGACCTCTCTCGACAAGCTCC 23563
Qy 1918 sCysProLeuGlyGlyThrLeu-----AspGI 1927
Db 23564 GGAGCCGCTCGAGAGCGAGCAGCTCCGCTCGCTCGCGCGCTGGTGTGGACGAGACGG 23623
Qy 1927 yValTyProThrLeuMetGluProValLeuLeuProLysGluAlaPro--ArgValAl 1946
Db 23624 CCCTGCTCTCGGCCACGAAGCGG-----CTTCCAGCTCGACCCCGACAGGCTTCT 23677
Qy 1946 aArgProGluArgProArgAlaAsp-----ThrGlyHisAlaPheLeuAlaLy 1962
Db 23678 TCGACCT--CGGTCTCGATTGATCATGACCGTCGAGCTTCGTGCGCGCTTGCAAC-AG 23733
Qy 1962 sProProAlaArgSerGlyLeuGluProAlaSerSer----- 1974
Db 23734 GCCACCGGCATCAAGCTCCCGGCCACCTCGCTTCGACCATCCCTCTCTCATCGCGTC 23793
Qy 1975 -----ProSer--LysGlySerGluProArg-- 1982
Db 23794 CGGCTTTCATCGCGGAGTCTCGCGGATGTCGACGCTCTTTGGGATTCCTCCACCAAGG 23973
Qy 1983 -----ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
Db 23854 ACGCGCGCGCTCGCGCGCGCTCGAGCGACGAGCCCATCGCTCGGATGGCC 23913
Qy 1998 o----- 1998
Db 23914 CTGGCGCTCGCGGCGGCTCGCGGATGTCGACGCTCTTTGGGATTCCTCCACCAAGG 23973
Qy 1999 -AlaLysAsnLeuAlaProHisHisAlaSerProAspProAlaPro----- 2014
Db 23974 CGCAGCGGTCGAGCCCATTCACAGAGCGCTGGAGCGCGCTGCGCTCTACGACCCC 24033
Qy 2014 ----- 2014
Db 24034 GACCCGACGCGGACGCCACAGAGCTAGCTCGGCATGCGCGATGCTCGACCATGAC 24093
Qy 2015 -----ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
Db 24094 CTCTTCGACCTGCTCTTCTCGGCATCAGCCCCCGGAGGCCAACACC---TCGACCCC 24150
Qy 2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerPr 2050
Db 24151 CAGCACCGCTGCTCTCTGATCTGCTGGCTGGCTCGACGCGCGGATGCTGCTCCC 24210
Qy 2050 oGluGlyValGluProValSerProValSerSer----- 2061
Db 24211 ACCTCCCTCAAGGACTCCCTCAGCGGCTCTCTGTCGCGATCTGCGCGGCGGAATACGG 24270
Qy 2062 -----ProSerLeuThrHisAspLysGlyLeuProLysHi 2073
Db 24271 ATCAAGAGGCGAGCTCGAAGGTTCCAGGTTTACTTTCATCCAAAGGACTTCG----- 24325
Qy 2073 sLeuGluLeuLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly-- 2092
Db 24326 -----GTCCTTTGGCGCGGGGGG 24345
Qy 2093 -----ProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGI 2111
Db 24346 TTGGCTATACGCTCGGCTCCAGGGCGG-----GATCTTCGTCGACACCGCC 24396
Qy 2111 uSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr 2131
Db 24397 TGCTCTCTCTCGCTCTCTCC-----TCCACCTCGCTCTCCCAAGCCCTCCGACAGGGC 24450

Qy 2131 gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyThrArgHisHi 2151
Db 24451 GAGTGCAACC----- 24460
Qy 2151 sProGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPhe-----Pr 2167
Db 24461 -----TCGCCCTCGCGCGGGGTGCTGCTCATGTTCTCCCCCAGACC 24504
Qy 2167 oGlyAlaSerCysProValLeuAspLeuArgArgProProProAspLeuTyLeuProPr 2187
Db 24505 TTGCTCATCTTTCCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCGCGCC-----GCTCCAG 24555
Qy 2187 oProAspHisGlyAlaProAla-----ArgGlySerProHisSerGI 2201
Db 24556 ACCTTCTCGACACCGCGAGCGCTACGAGCGGAGAGGCTGCTGCTTCCTTCGCCTC 24615
Qy 2201 uGlyGlySerArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGI 2221
Db 24616 GAGCGGATCGCGACGCGCTTCGCGCGGAGACACCGG-----TCCTC 24657
Qy 2221 yIleGluProValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaVa 2241
Db 24658 GTCTCTGTCGCGCGACCGCCATCAACACGAG-----GCGGTCGAGCGGT 24705
Qy 2241 lTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe 2261
Db 24706 ATCACCG-----CCCCAACGCGACCTCCACGAGA-----AGGTC 24741
Qy 2261 rProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAl 2281
Db 24742 CTCGGGCGCGCTCCACGAGCGCGCATCCCGCGCGACGCTCGAGCTCGTGCAGTGC 24801
Qy 2281 aMetValLysSerLysLysGlnGluIleAsn----- 2291
Db 24802 CATGCAACCGGCACTCTCGGGAGACCCCATCGAGTCAAGCCCTGCGCGCGCTCTAC 24861
Qy 2292 -----LysLysLeuAsnThrHisAsnArgAsnGluProGluTyAs 2305
Db 24862 GCCGACGCGACACCGCTGAAAGCTCTCTCTCGCGCGCTCAAGACCAACATCGGC 24921
Qy 2305 nIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMe 2325
Db 24922 CATCTCGAGCGG-----CTCGCGGCTC-GCGGCGCTCGCCAAGAT 24962
Qy 2325 tThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaI 2345
Db 24963 GGTG-----GCCTCGCTCGCGACGCGCTGCCCGCCACCTCCACGCGAC 25010
Qy 2345 eIleArgLysAlaLeuMetGlyLysTyAspGlnTrpGlu-GluSerProLeuSerA 2365
Db 25011 CCCACGCAATCCCTCATC-----GAGTGGAGCGCTCGCCATCGACGCTCGT 25058
Qy 2365 laAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrA 2385
Db 25059 CG-----ATACCCCGAGCGCTTGGCCCGCCACGAGATGGCGAGTCCCCCGCGC 25109
Qy 2385 laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLys----- 2402
Db 25110 CGGCATCTCGCGCTTCGATTCCTCGGCACCAACGCGCCATCTCTCGAAGGGTCC 25169
Qy 2403 -----AlaLysValSerGlyArgProSerArgLysAlaLysSer---ProAlap 2419
Db 25170 CGCGCGCTCGCGCGCGCGCCACCTCACAGCGCGCTCGCAACGCGCTCCCGCGGC 25229
Qy 2419 roGlyLeuAlaSer-----GlyAspArgProProSerValSerValHisSerG 2436
Db 25230 GTGGCCGCTGCTCTGTCGCGCGAGGAGCGCGCGCG-----TCGCGCGCCAGGCGAA 25283
Qy 2436 luGlyAspCysAsnArgArgThrProLeuThr-----AsnArgValTrpGluAspA 2453
Db 25284 CGGCTCGCGGACCACTCTGCTCGCCACGACGACCTCACCTCGCGATGTGGCTATTTC 25343
Qy 2453 rgProSerSerAlaGlySerThrProPheProTyAsnProLeuIleMetArgLeuGlnA 2473

Db 21343 GCGCCCTGCTCGCGATCTGGAGTCCGAGGGGCTCTTCCGCCCTCAAGCTGAGTACGAC 21402
Qy 1353 -----GlySerIle---ThrGlnGlyIleProArgSerTy-ValG 1365
Db 21403 TTGCTCTCCACCTCCGCGCAGGTGAGTGCATTCGGCAGAGCTCTCGATCTCTGTCG 21462
Qy 1365 luAlaGln-----GluAspTyLeuArgA 1373
Db 21463 TGGCTCGAGCGCGCTCGAGCGGGTCCCGTTCTACTCCACGGTGTAGCGCGCGCATC 21522
Qy 1373 rgGluAlaLysLeuLysArgGluGlyThrProProPro----- 1387
Db 21523 GACGGGAGCGAGCTCGACGCGCGCTACTGTGTACCGGAACCTCCGCGAGCGGTCCGCTTC 21582
Qy 1388 -----ProProSerArg----- 1391
Db 21583 GCAGACGCTGTGCAAGCGCTCTTGGCGGAGAACATCGCTTCTGTGGAGGTGAGCCCC 21642
Qy 1392 -----AspLeuThrGluAlaTyLeuLysThrGlnAlaLeuGlyProLeu----- 1405
Db 21643 AGTCTGTGTGACCTTGGCTTGGCAGAGCTCTCGAAGCGTGGAGCGCTCGCGGGCG 21702
Qy 1406 -----LysLeuLys-----ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
Db 21703 GTGCTCGGCTCTCTGTGGAGCGACGAAGGGGATCTACGGGGCTTCTCGTCTCGCTCTCC 21762
Qy 1423 rgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlap 1443
Db 21763 GAGCTTACGTCAACGGCTTCGCCCTGGATGGACGACATCTT-----GCC 21810
Qy 1443 roArg-----ProLeuLysGluGlySerIleThrGlnGlyT 1455
Db 21811 CCGGGAAGCGGTGCGCTGCCACCTACCCCTTCC-AGCGGAGCGCTTCTGGCTCGA 21869
Qy 1455 hrProLeuLysTyAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgS 1475
Db 21870 CGCTTCCAGCGCACCGCCCGCGCGCTCAACCCACTTGTCTCCGCTCG-AGGG 21920
Qy 1475 erLeuIleGlySerProGlyArg----- 1482
Db 21921 GCGGTTCTGCGAGCCATCAGAGCGGGAATATCGAGCGCGCTCAGCGGCCAGCTCCACGT 21980
Qy 1483 -----ThrPheProValHisProLeuA 1491
Db 21981 GGAGCGGACGAGCAGCGCGCCCTTGCCTTCCCTTCCACCCCTCGCGAGCTTTCG 22040
Qy 1491 spValMetAlaAspAlaArgAlaLeu-GluArgAlaCysTyThrGluGluSerLeuLysSer 1510
Db 22041 CCACGAGCGGCAAGAGCAGCGGCGTGCAGCGCTCGCGCTACCGCATCAGTGGAGGCC 22100
Qy 1511 -----AtgProGly-----ThrAlaSerSerSerGly 1519
Db 22101 TCTGACACCGCCACACGCGCGCGGACCTGGCGGACCTGGCTCTCTGTCGCGCGC 22160
Qy 1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro-----Arg 1537
Db 22161 CGCTTGGACGACGACGCGCTCCCTCCGCGCTCACCAGGCGCTCGCCCGCGCGCGCGC 22220
Qy 1538 GlnSerProLeuThrTyThrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly 1557
Db 22221 GCGGCTCTCGCGCTGCGCTGAGCCAGGCCACCTTGA-----CCGCGAGGC 22268
Qy 1558 SerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSer 1577
Db 22269 TCT-----CGCCGAGCACCTGGCGCAGGCTTGGCGGAGACCGCGCGCTCGCG 22319
Qy 1578 LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis 1597
Db 22320 CGTGTCTCGCTCTCGCTCGACGAAAGTCCCTCGCGACCATCGCCGCGCTGCGCGC 22379
Qy 1598 SerThrValProGluHisHisProHisProIleSerProTyThrGluHisLeuLeuArgGly 1617

22380 GGGACTCGGCTTCTCGCTCACCTCTCGTCCA---AGCCCTCGGCGCATCGCCCTCGACGC 22436
Qy 1618 ValSerGlyValAspLeuTy-----ArgSerHisIle 1628
Db 22437 GGCCTT---GTGGCTTTCACCGCGCGCGCTCTCGTTCGGACACTCCGACCCCATCGC 22493
Qy 1629 Pro-----LeuAlaPheAspProThrSerIleProArgGlyIlePro 1642
Db 22494 CCATCGACGCGAGCGCATGACCTGGGGCTGGGGCGCGCTCGTTCGGCTTCGAGCA---CCC 22550
Qy 1643 LeuAspAlaAlaAlaAlaTyTyLeuProArgHis----- 1654
Db 22551 CGAGCCCTGGGAGGGCTCTCGACCTCGGCGGAGCATCGACCGAGCCCTGGGGCGC 22610
Qy 1655 LeuAlaProAsnProThr-----TyProHisLeuTyProPro---Tyr 1668
Db 22611 CTGTCTCCCGTCTCGCCCTCGCAACGATGAGGACAGCTCGCTCTCGCCCGCGCGC 22670
Qy 1669 LeuIleArgGlyTyProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
Db 22671 GTTCTACGCTCGCGCTCTCGCTCGCGCTCGCTCGCGAGCGCGCGC-----CGCAGC 22724
Qy 1689 TyrIleThrSerGlnGln-----MetHisHisAsnThrAlaThrAlaMetAlaGlnArg 1706
Db 22725 TACCTTCAAGCCCGAGGACCCCTCTCTATCACCAGAG-GCACCGGCGCGCTGGCGCTC 22783
Qy 1707 AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyAlaAla 1726
Db 22784 ACGTCCCGCATGCTCGCTCGAGAGGCGCAGACGACCTCGTCTCATCAGCGCGCGAG 22843
Qy 1727 GlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
Db 22844 GGGCCAGGCGGAGGCGCTCGAGCTCCACCGCGAGCTCAGCGCTCGGGCGCGCGC 22903
Qy 1747 ThrProGly-----ThrProAlaThrAlaMetAspArgLeuAlaTyLeuPro 1762
Db 22904 TCACCTTCGCGCGGTGTGTGTCGCGCAGAGGAGCGCTGTGCCACGCTTCTCGACGAGC 22963
Qy 1763 ThrAlaProGln-----ProPheSerSerArgHisSerSerSer----- 1775
Db 22964 TCGACCGCGAAGGTCGAGGTCGCGCGCTGTTCACGCGGCGCGCATCGGGCGCCACG 23023
Qy 1776 -----ProLeuSerProGlyGlyProThr----- 1783
Db 23024 CTCGCTCGCGCGCACCTCTCTCATGAGCTCGCGCGAGTGTCTCTGCCAAGTCTCTAG 23083
Qy 1784 -----HisLeuThrLysProThrThrThrThrThrThrThrThrThrThr 1795
Db 23084 GCGCAGGAACCTCCACGACCTCGTCTCGTCTCGACCCCTCGACGCGCTTCTCTTTTCT 23143
Qy 1796 Arg-GluArgAspArgAspArgAspArgAspArgAspArgAspArgAspArgAspArg 1815
Db 23144 CGTCCATCGAGCGCTGCGGGCGCGGACACAAAGCGGATACGCGCGCGGAAACGCT 23203
Qy 1815 uThrSerThrThrThrValGluHisAlaProIleTrp----- 1827
Db 23204 TCCTCGACGCGCTGGCGGACCGACGCGCGAGTCTTGACAGCCGAGACACGTCCGTGTGT 23263
Qy 1828 -----ArgProGlyThrGluGlnSerSe 1835
Db 23264 GGGCGCGTGGGCGCGCGGTGTATATTACAGGGGCGCTTGGACCGCCAGCTGAGC 23323
Qy 1835 rgLysSerSerGlySerSerGly-----GlyGlyGlyGlySerSerSerArgProAl 1852
Db 23324 AACGTCGTCTGTGCGCGATGCGCCCTTTCGTGGCGGTGGCGGCTCGCGCAAGCCCTGG 23383
Qy 1852 aser-----HisSerHisAlaHisGlnHis----- 1860
Db 23384 AGCAGCAGGAGACACCGCTCACGCTCGCGGACATCGACTGGGCGCGCTTTGCGCTTCGA 23443
Qy 1861 -----SerProIleSerProArgThrGlnAspAlaLeuGlnGlnProSerValLe 1878
Db 23444 TCAGCGTCTCGCTCCCGCGCTCTCTCGCGGAGTTCGCCCGGAGCGC----- 23492

| | | | |
|----|-------|---|-------|
| QY | 731 | lProArg-----GlyGluCysSerGlyProAlaThrVa | 742 |
| Db | 19357 | GCCTCTACGACCCGACCCGACGCAAGCCCAAGAGCTACGTCGGCATCGCGCATG | 19416 |
| QY | 742 | lAsnAsnSerSerAspThrGluSerIleProSer-----ProHi | 755 |
| Db | 19417 | CTCGACCAAGTCGACCTCTTCGATCTGCTGCTTTTGGCATCAGCCCTCGCGAGGCCAA | 19476 |
| QY | 755 | sThrGluAlaAlaLysAspThr---GlyGlnAsnGlyProLysProAlaThrLeuGl | 774 |
| Db | 19477 | TACCTCAGCCCCAGCAGCCGCTCTCTGATCTGCTGGCTGGCCCTCGAGGAGCC | 19536 |
| QY | 774 | yAlaAspGlyProProGly-----ProProThr----- | 784 |
| Db | 19537 | GGCATGCTCCCTCCACCTCAAGGATCTCCACCGCGCTCTTCGTCGGCATCGCGCC | 19596 |
| QY | 785 | -----ProProArgArgThrSerArgAlaProIleGluPr | 796 |
| Db | 19597 | AGCGAATACGACCTCGAAACACAGAGCTCCGAGAGGTGGAAGGTATGCTCTCCAGGC | 19656 |
| QY | 796 | oThrProAlaSerGluAlaThrGlyAla---ProThr----- | 807 |
| Db | 19657 | ACCCCGGGTCTTTGCGGGGGCGCTTGGCTTACACAGCTCGGCCTCGCAAGGCGCCGCG | 19716 |
| QY | 808 | -----ProProAlaProProSerProSerAlaProProValValProLy | 824 |
| Db | 19717 | CTCTCGGTGCGACACCGCTGCTCTCTCTGCTGCTGCG-----CCCTCCACCTCGCCTGC | 19770 |
| QY | 824 | sGluGluLysGluGluThrAlaAlaAlaProProValGluGluGluGlnLy | 844 |
| Db | 19771 | CAAGCCCTCCAGAGGGGAGTGCACTCCCTCGCCCTCGCCGGGGCTCTCGTCATGGCC | 19830 |
| QY | 844 | sProProAlaAlaGluGlu-----LeuAlaValAs | 854 |
| Db | 19831 | TCCCGGGCTCTTCTGCTCTTTCGCGATGCTGCTTTGGCGCCCGATGGCGCTCC | 19890 |
| QY | 854 | pThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGlyPr | 874 |
| Db | 19891 | AAGACCTTCTCGACCAACCGAGCGGTACCGAGCGGAGAGGGCTGCTGCTCTGCTGCC | 19950 |
| QY | 874 | oAlaLysGlyLysAspAlaGluAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGl | 894 |
| Db | 19951 | CTCGAGCGCTCGCGAGCGCTCGCCGAGGACACCGCGCTCTGCGCTCGTCGCGGC | 20010 |
| QY | 894 | uLysLysGluGlySerGlyArgAlaThrThralalysSerSerGlyAlaProGlnAs | 914 |
| Db | 20011 | ACCGCATGAACATGACGGCGCTGAGCGGCATCACCGCCCGCAATGGCACCTCCAC | 20070 |
| QY | 914 | pSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyAspLy | 934 |
| Db | 20071 | CAGAGGTCTCTCGCGCGCGCTCCAGCGCCATATCGGCCCTCGCGAGCTCGACGTC | 20130 |
| QY | 934 | sAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAs | 954 |
| Db | 20131 | GTCGAATGCCATGGCAGCGCACCTCTTGGGAGACCCCATCG-----AGGTGCAA | 20181 |
| QY | 954 | nAla---SerProGlnLysProLeu-----AspLeuLysGlnLeuLysGlnAlaAl | 971 |
| Db | 20182 | GCCCTGGCGCGCTTACGCCGATGGCAGACCCGCTGAAAGCCCTCTCTCTCGCGCA | 20241 |
| QY | 971 | aAlaIleProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaPr | 991 |
| Db | 20242 | CTCAGACCAACATTGGCCATCTCGAG-----CGCCCTCGGCTCGCGGCGCTGCC | 20295 |
| QY | 991 | oThrLysProAlaPro-----ProAlaProProPro-----ProGlnAs | 1004 |
| Db | 20296 | AAGATCGTCCCTCTCGCCCATGACGCGCTGCCCCCACCCTCCACACGACCCCGCGC | 20355 |
| QY | 1004 | nLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer-----ProArgGlyLy | 1022 |
| Db | 20356 | AATC---CCCTGATGATGGATGGCTCGCATCGAGCTGCTCGATCCACAGAGGGG | 20412 |
| QY | 1022 | sSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe | 1042 |
| Db | 20413 | TGGG-----CCCGCCACAGATGSCAGTCCCGCGCGCGCGCGCTCTCC | 20457 |
| QY | 1042 | uProGlyAspProProCysTrpThrSerGlyLeuProPhe----- | 1055 |
| Db | 20458 | GCCTTCGGACTCTCGGACCAACCGCCACCGTTATCTCGAAGAGGCTCCCGCGATCCCG | 20517 |
| QY | 1056 | -----ProValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAl | 1074 |
| Db | 20518 | CAGGCCGAGCCACCGCGGCACAGCTCGCTCGCAGC---CGCTCCCGCAGCCTGGCCC | 20574 |
| QY | 1074 | aPheSerTyraAlaPro-----ProGlyHisProLeuProLeuGlyLeuHisAs | 1090 |
| Db | 20575 | GTGCTCTCTCGCCAGGAGCGCGCGCTCGCGCCAGGCCAGAGGCTCGCGCAG | 20634 |
| QY | 1090 | pThrAlaArgProValLeuProArgPro-----ProThrIleSerAsnProPr | 1106 |
| Db | 20635 | CACCTCTCGCCACGACACCTCGCTTGGCCGATGAGCTACTCTGCTCGCCACACCC | 20694 |
| QY | 1106 | o-----ProLeuIleSerSerAlaLysHisProSerValLeuGl | 1119 |
| Db | 20695 | CGGGCTACCTTCGAGCACCGCTCGCTCTCGTGTCCAGCGCGAGAGCTCTCTCTCC | 20754 |
| QY | 1119 | uArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTySerGl | 1139 |
| Db | 20755 | GGCTCGATTGCTCGCCAGGAGAGCGCGCC-----CGAGCACCGCTCGTCGAA | 20805 |
| QY | 1139 | uHisAlaLysAlaPro-----ValGlyPro-ValThrMetGlyL | 1152 |
| Db | 20806 | CGAAGCGGAGCACCGCAAGGTGCTGCTTCTTCTGGGCAAGGCTCGCAGTGGAA | 20865 |
| QY | 1152 | euProLeuProMetAspProLysLysLeuAlaProPheSerGly---ValLysGlnGluG | 1171 |
| Db | 20866 | GGATGGCTCTCTCTGCTCGATACCTCGCGGTCTCGGCGCACAGCTCGAAGCGTGC | 20925 |
| QY | 1171 | lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG | 1191 |
| Db | 20926 | GAGCGCGCTCGCGCCCGCCACGTGAGTGTGCTGCTCGCGTCTCGCGTCTCGCGG | 20978 |
| QY | 1191 | luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyLysIleThrLysG | 1211 |
| Db | 20979 | -----CGAGGA-----GGGCGCGCCCGCTCGACCGGTCGACGTG | 21015 |
| QY | 1211 | lyIleProSerThrArgValProSerAspSerAlaIleThrTyArgGlySerIleThrH | 1231 |
| Db | 21016 | GT-----CCAGCCCGCTGTTCTCGATGATGCTCTCGTGGCGCGCTGTGCGCTCC | 21069 |
| QY | 1231 | isGlyThrProAlaAspValLeuTyLysGlyThrIleThrArgIleIleGlyGluAspS | 1251 |
| Db | 21070 | ATGGCGCTCGAGCC-----CGACGGGTGGTGGCCAT | 21102 |
| QY | 1251 | erProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrg | 1271 |
| Db | 21103 | AGCAGCGCGAGATCGCGCGGC-----CTGTGCGCGCGCGCTGCTGCTC | 21150 |
| QY | 1271 | luGlyLysLysGlyHisValLeuSerTyrgLuglyGlyMetSerValThrGlnCysSerL | 1291 |
| Db | 21151 | GAGGACGC-----TGCCAAGCTGGTGGCTGCGCGACCGCGCGCTC | 21192 |
| QY | 1291 | ys-----GluAspGlyArgSerSerS | 1298 |
| Db | 21193 | GTGAGCTCGCCCGCAGGGGGCCATGCGCGGTGGAGTGGCGGAGCGCGAGTCCGCA | 21252 |
| QY | 1298 | erGlyProProHisGluThrAlaAlaProLysArgThrTyArgMetMetGluGlyArgv | 1318 |
| Db | 21253 | CGCGCTCTCA-----GCCTATGGCGATCGGCTC | 21282 |
| QY | 1318 | alGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProG | 1338 |
| Db | 21283 | TCCATCGGGCGGATCAACAGCCCTCGTTTACGACGATCTCCGCGAGCCCTCGCGTC | 21342 |
| QY | 1338 | luArgHisSerProHis-----HisLeuLysGluGlnHisIleArg- | 1352 |

| | | | | |
|--------|---|---|----|---|
| DB | 24406 | ATCACCG-----CCCCCAACGGCACCTCCACGACGAG-----AGTCC 24441 | OS | Polyangium cellulorum. |
| QY | 2261 | rProGlyAanThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAl 2281 | PN | US5662898-A. |
| DB | 24442 | CTCCGGCGCGCTCCACAGACCCCGCATCACCCCGCGGCGTGGAGTGC 24501 | XX | 02-SEP-1997. |
| QY | 2281 | aMetVallyserLysGlnGluLeuAsn----- 2291 | XX | 01-JUN-1995; 95US-00457342. |
| DB | 24502 | CATGGCACCGACCTCGTGGGAGACCCCATGAGGTGCAAGCCTGGCGCGCTTAC 24561 | PR | 20-AUG-1990; 90US-00570184. |
| QY | 2292 | -----LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAs 2305 | PR | 02-JUL-1992; 92US-00908284. |
| DB | 24562 | GCGACGGCAGACCCGCTGAAAGCTCTCTTCGCGCGCTCAAGACCAATCGGC 24621 | PR | 31-AUG-1992; 92US-00937648. |
| QY | 2305 | nileSerGlnProGlyThrGluilePheAsnMetProAlaileThrGlyLeuMe 2325 | PR | 01-JUL-1993; 93US-00087636. |
| DB | 24622 | CATCTCGAGCCG-----CCTCGGGCTC-GCGGGCGTGCACAGAT 24662 | PR | 08-JUN-1994; 94US-00258261. |
| QY | 2325 | tThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluail 2345 | PA | (CIBA) CIBA GEIGY CORP. |
| DB | 24663 | GGTC-----GCCTCGCTCCGCCACGAGCCCTGCCCCCACCCTCCAGCGAC 24710 | XX | Ligon JM, Hill DS, Lam ST, Hammer PE; |
| QY | 2345 | eileAtGLysAlaLeuMetGlyLysTyrAspGlnGlu-GluSerProProLeuSerA 2365 | XX | WPI; 1997-447901/41. |
| DB | 24711 | CCACGCAATCCCTCATC-----GAGTGGAGGCGCTCGCCATCGAGTGGT 24758 | DR | Protecting plants against pathogens with genetically transformed |
| QY | 2365 | laAnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrA 2385 | PT | biological control agent - which expresses all polypeptide(s) involved in |
| DB | 24759 | CG-----ATACCCCGAGGCGCTTGGCCCCCGACGAGATGGCAGTCCCCCGCGCGC 24809 | PT | pyroclitrin biosynthetic pathway. |
| QY | 2385 | laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLys----- 2402 | XX | Example 16; Col 104-128; 88pp; English. |
| DB | 24810 | CGGCATCTCCGCTTCGATTCCTCGGCACCAAGCCACGTCATCTCGAAGAGGCTCC 24869 | CC | This genomic DNA sequence encodes the Soraphen gene cluster. This |
| QY | 2403 | -----AlaLysValSerGlyArgProSerSerArgLysAlaLysSer---ProAlap 2419 | CC | sequence encodes two open reading frames (ORF's), the positions of which |
| DB | 24870 | CGCGCGCTCGCGCGGCGCGCCACCTCACAGCGCGGTGCGAAGCGCGCTCCCGCGC 24929 | CC | are not given in the specification. ORF1 is approximately 25.5 kb in size |
| QY | 2419 | roGlyLeuAlaSer-----GlyAspArgProProSerValSerSerValHisSerG 2436 | CC | and encodes 5 biosynthetic modules homologous to the erythromycin genes |
| DB | 24930 | GTGGCCCGTGTCTGTGCGGACGAGGCGGCGCG-----TCGCGCGCCAGCGCAA 24983 | CC | of Saccharopolyspora erythraea. Each module contains a beta- |
| QY | 2436 | luGlyAspCysAsnArgThrProLeuThr-----AsnArgValThrGluAspA 2453 | CC | ketoacylsynthase (KS), an acyltransferase (AT), a ketoreductase (KR) and |
| DB | 24984 | GCGGCTCGCGCACCACTCGTGGCCGACGAGCCTCACCTCGCGATGTGGCTATTTC 25043 | CC | an acyl carrier protein (ACP) domain. ORF2 is immediately adjacent to |
| QY | 2453 | rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuileMetArgLeuGlnA 2473 | CC | ORF1 and is thought to contain 3 modules. Soraphen is a type I polyketide |
| DB | 25044 | GC----- 25045 | CC | synthase (PKS) which has antibiotic activity. Transgenic plants |
| QY | 2473 | laGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaG 2493 | CC | containing such antipathogenic genes like those encoded in the Soraphen |
| DB | 25046 | -----TGGCCACCAACCGCGCCACTTCGACGACCGCGCGCTCTCGT 25088 | CC | cluster should have enhanced resistance to attack by phytopathogens. |
| QY | 2493 | lyProHisHisAlaTrpAspGluGluProLys 2503 | CC | (Updated on 14-MAY-2003 to correct PS field.) (Updated on 17-OCT-2003 to |
| DB | 25089 | AGCCCAACACCGCGAGAGTCTCTCTCGCGC 25120 | CC | standardise OS field) |
| RESULT | 54 | | XX | Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 U; 0 Other; |
| ID | AAT89956 | | XX | |
| ID | AAT89956 | | XX | |
| AC | AAT89956; | | XX | |
| XX | | | XX | |
| XX | | | XX | |
| DT | 17-OCT-2003 (revised) | | XX | |
| DT | 14-MAY-2003 (revised) | | XX | |
| DT | 12-MAR-1998 (first entry) | | XX | |
| DE | Sorarium cellulorum soraphen gene cluster genomic DNA. | | XX | |
| XX | | | XX | |
| XX | Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic; | | XX | |
| XX | transgenic plant; phytopathogen; resistance; ss. | | XX | |
| XX | | | XX | |

```
QY 1707 AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAenTyAlaAla 1726
DB 22484 ACCTGCCCGGATCGCTCGTTCGAGAGGCGCAGAGCATCGTCTCATCAGCGCGGAG 22543
QY 1727 GlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
DB 22544 GGCCTCCAGGCGGAGGCGCTCGAGGCTCCACCGCGAGCTCACGGCCCTGCGGCGCGCGG 22603
QY 1747 ThrProGly-----ThrProAlaThrAlaMetAspArgLeuAlaTyLeuPro 1762
DB 22604 TCACCTTCGCGCGGTGTGATGTCGCCGACAGAGCGCTGTCGCCACGCTTCTCGAGCAGC 22663
QY 1763 ThrAlaProGln-----ProPheSerSerArgHisSerSerSer----- 1775
DB 22664 TCGACGCCGAAGGTCGACGTCGCGCGCGGTGTTCACGCGGCGGCGCATCGCGGCGCACG 22723
QY 1776 -----ProLeuSerProGlyGlyProThr----- 1783
DB 22724 CTCGGCTCGCGCGCATCTCTCATGAGAGTCGCCGACGTTGTCTGCCAAGTCTTAG 22783
QY 1784 -----HisLeuThrLysProThrThrThrThrSerSerSerGlu 1795
DB 22784 GCGCAGGGAACCTCCAGCACTCTCGTCTCGACCCCTCGAGCGCTTGTCTCTTTTCT 22843
QY 1796 Arg-GluArgAspArgAspArgGluArgAspArgAspArgGluArgGlySerIleLe 1815
DB 22844 CGTCCATCGAGCGCTGCGGCGGCGGACAAAGCGGATACGCGCGCGGAACGCCT 22903
QY 1815 uThrSerThrThrValGluHisAlaProIleTpo----- 1827
DB 22904 TCTCGACGCCCTCGCGGACCAAGCGCGCAGTCTTGACACGCGGACACGCTCGTGGTGT 22963
QY 1828 -----ArgProGlyThrGluGlnSerSe 1835
DB 22964 GCGCGCGTGGGCGGCGGTGTATATTCACGGGCGCTTGCAGCCACCTCGAGC 23023
QY 1835 rGlySerSerGlySerSerGly-----GlyGlyGlySerSerSerArgProAl 1852
DB 23024 AAGCTCGTCTGTCGCGATGGCCCTTTCGTCGCGCGTGGCGGCTCGCGCAAGCCCTGG 23083
QY 1852 aSer-----HisSerHisAlaHisGlnHis----- 1860
DB 23084 AGCACGACGACACACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTGGCGCTTGA 23143
QY 1861 -----SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValle 1878
DB 23144 TCAGCGTCGCTCGCTCCGCGCGCTCTCGCGACTTCCCGAGCAGCGC----- 23192
QY 1878 uHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValle 1898
DB 23193 -----GCCCTCGAAGAC----- 23204
QY 1898 uArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHi 1918
DB 23205 -AGAGAGGCGCTCTCTCTCCGAGCAGCGCGCGCGCGCCCGACCTCTCTCGACAAGCTCC 23263
QY 1918 sCysProLeuGlyGlyThrLeu-----AspG1 1927
DB 23264 GGAGCGCTCGAGAGCAGCAGCTCGCTGTCTCGCGCGCTGTGTGCGAGCAGCGG 23323
QY 1927 yValTyProThrLeuMetGluProValLeuLeuProLysGluAlaPro---ArgValAl 1946
DB 23324 CCCTCGTCTCGGCCACGAAGGCG-----CTTCCAGCTCGACCCCGACAAAGGCTTCT 23377
QY 1946 aArgProGluArgProArgAlaAsp-----ThrGlyHisAlaPheLeuAlaTy 1962
DB 23378 TCACACT---CGGTCTCGATTCTCATCATGACCTGAGCTTCTGCGCGCTTCTGCAAC-AG 23433
QY 1962 sProProAlaArgSerGlyLeuGluProAlaSerSer----- 1974
DB 23434 GCCACCGGCATCAAGTCCCGGCCACCTCGCTTCGACCATCCTCTCTCATCGCGTC 23493
QY 1975 -----ProSer---LysGlySerGluProArg-- 1982
```

```
DB 23494 GCCTCTTCATGCGCGCACTCGCTCGCCCGCAGCGCTCGGACAGGCTCTCCGCGAGCG 23553
QY 1983 -----ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
DB 23554 AGCCGCCCGGCTCGCGCGCTCGAGGAGGCCCATCGCCATCGTCGCGCATGGCC 23613
QY 1998 o----- 1998
DB 23614 CTGCGCTCGCGGCGCGCTCGCGCATGTCGACGCTCTTTGGGAGTTCTCTCCACCAAGGG 23673
QY 1999 -AlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaPro----- 2014
DB 23674 CCGCAGCGCTCGAGCCCATTCACAGACGCGCTGGGACCGCGTCTTACGACCCC 23733
QY 2014 ----- 2014
DB 23734 GACCCGACGCGCAGCGCAAGACTAGTCCGGCATCGCGCATGCTCGACCATCGAC 23793
QY 2015 -----ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
DB 23794 CTCTTCGACCCCTGCTTCTTCGGCATCAGCCCGGGAGGCCAACACAC---TCGACCCC 23850
QY 2030 oPheSerIleGlnGluLeuLysLeuArgSerLeuGlyTyHisGlySerSerTySerPr 2050
DB 23851 CAGCACCGCTGCTCTCGAATCTGCTGCTGCGCTCGAGCACCGCGCATCGTCCCC 23910
QY 2050 oGluGlyValGluProValSerSer----- 2061
DB 23911 ACCTCCTCAAGACTCTCTCACCGCGCTCTGCTCGGCATCTGCGCGCGGAATACGG 23970
QY 2062 -----ProSerLeuThrHisAspLysGlyLeuProLysHi 2073
DB 23971 ATGCAAGAGCGAGCTCGGAAGTTCCGAGTTTACTTATCCAAAGCACTTCG----- 24025
QY 2073 sleuGluGluLeuAspLysSerHisleuGluGlyGluLeuArgProLysGlnProGly-- 2092
DB 24026 -----CGTCTTTGGCGCGGGGGC 24045
QY 2093 ----ProValLysLeuGlyGlyAlaAlaHisLeuProHisLeuArgProLeuProGl 2111
DB 24046 TTGGCTATACGTCGGCTCCAGGGCGCG-----GATCTTCGGTCGACACCGCC 24096
QY 2111 uSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValysGlyHisGlnAr 2131
DB 24097 TGCTCTCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 24150
QY 2131 gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHi 2151
DB 24151 GAGTGCAACC----- 24160
QY 2151 sProGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPhe-----Pr 2167
DB 24161 -----TCGCCCTCGCGCGGGCGTGTGCTCATGTCTCTCTCTCTCTCTCTCTCT 24204
QY 2167 oGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyLeuProPr 2187
DB 24205 TTGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 24255
QY 2187 oProAspHisGlyAlaProAla-----ArgGlySerProHisSerGl 2201
DB 24256 ACCTTCTCGACACACCCCGACGCGCTACGACCGCGGAGAGCGTGTCTCTCTCTCTCTCTCT 24315
QY 2201 uGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGl 2221
DB 24316 GACCGGATCGCGCAGCGCTCTCGCGCGGAGACACCGG-----TCCTC 24357
QY 2221 yIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVa 2241
DB 24358 GTCTCTCGTCGCGGACCGGCATCAACACGAGC-----CGCGGTCTGAGCGGT 24405
QY 2241 lTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe 2261
```


| | | | | | | |
|----|-------|--|-------|-------|--|-------|
| Db | 20455 | GCCTCGATTTCGTCGCCAGGGAAGGCCGCCCC-----CGAGCACCCTGCTCGTAA | 20505 | 1423 | rgSerIleHisGluLeuProArgGluGluLeuArgHisThrProGluLeuProLeuAlaP | 1443 |
| Qy | 1139 | uHisAlaLysAlaPro-----ValGlyPro-ValThrMetGlyL | 1152 | 21463 | GAGCTCTACGTCAACGGCTTCGCCCTGGATTGGACGATCCT-----GCC | 21510 |
| Db | 20506 | CGAAGCGGAAGCACCGCAAGGTCTTCTTCTTCTGGCAAGCTCGCAGTGGAA | 20565 | 1443 | roArg-----ProLeuLysGluGlySerIleThrGlnGlyT | 1455 |
| Qy | 1152 | euProLeuProMetAspProLysLysLeuAlaProPheSerGly-----ValLysGlnGluG | 1171 | 21511 | CCCGGAGCGGGTGCCTGCCCTACCTACCTTCC-AGCGCAGCGCTTCTGGCTCGA | 21569 |
| Db | 20566 | GGGATGCGCTCTCTCCCTGCTCGATACCTCGCGGTCTTCGGGACACAGCTCGAAGGTGC | 20625 | 1455 | hrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgS | 1475 |
| Qy | 1171 | lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG | 1191 | 21570 | CGCTCCAGCGACCGCGCGCGGTCAACACCTTGTCTCGCTCG-----AGGG | 21620 |
| Db | 20626 | GAGCGCGCTCGCGCGCCACCGTGAAGTGTGCTCGCGGTCTCCCGG----- | 20678 | 1475 | erLeuIleGlySerProGlyArg----- | 1482 |
| Qy | 1191 | luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG | 1211 | 21621 | GGGTTCCTGGCAGGCATCGAGCGGGNATATCGACGGCTCAGCGGCACCTCCACCT | 21680 |
| Db | 20679 | -----CGAGGA-----GGCGCGCCCGCTCGACCGCGGTGCGAGTG | 20715 | 1483 | -----ThrPheProProValHisProLeuA | 1491 |
| Qy | 1211 | lylleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrH | 1231 | 21681 | GGAGCGCAGCAGCAGCGCGCGCTTGCCTTCTTCTTCCACCTCGCGAGCTTTCG | 21740 |
| Db | 20716 | GT-----CCAGCGCGCGCTTCTCGATGATGCTCTCGTGGCGCGCTGTGGCGCTCC | 20769 | 1491 | spValMetAlaAspAlaArgAlaLeu-GluArgAlaCysTyrGluGluSerLeuLysSer | 1510 |
| Qy | 1231 | isGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspS | 1251 | 21741 | CCAGCAGCGCAAGACGAGCGGACGGTCCAGCGCTTCCCTACCGCATCAGCTGGAAGCC | 21800 |
| Db | 20770 | ATGGCGCTCGAGCC-----CGAGCGGTGTGCGGCAT | 20802 | 1511 | -----ArgProGly-----ThrAlaSerSerSerGly | 1519 |
| Qy | 1251 | erProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrG | 1271 | 21801 | TCTGACCAACCGCCACCGCGCGCTGCGCGCACCTGGCTCTCTGCTGCTGCGCG | 21860 |
| Db | 20803 | AGCCAGCGCGATCGCGCGGC-----CTGTGTGGCGCGCGCTGTGCTC | 20850 | 1520 | GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro-----Arg | 1537 |
| Qy | 1271 | luGlyLysLysGlyHisValLeuSerTyrGluGlyMetSerValThrGlnCysSerL | 1291 | 21861 | CGCTCTGGAGCAGCAGCGCTCCCTCCCGCTCACCGAGCGCTCGCGCGCGCGCGC | 21920 |
| Db | 20851 | GAGGACGC-----TGCCAAAGTGTGTGGCGCTCGCGCAGCGCTGCGCTC | 20892 | 1538 | GlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly | 1557 |
| Qy | 1291 | ys-----GluAspGlyArgSerSers | 1298 | 21921 | GGCGTCTCGCGTGGCGCTGAGCCAGCGCCACCTGGA-----CGCGAGGC | 21968 |
| Db | 20893 | GTGGAGCTCGCGCGCGCGCGCCATGCCCGGTGGAGTCCCGGAGCGCGAGTTCGA | 20952 | 1558 | SerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSer | 1577 |
| Qy | 1298 | erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV | 1318 | 21969 | TCT-----CGCGCAGCACCTCGCGCGAGGCTTGGCGAGACCGCGCGCTCGCGG | 22019 |
| Db | 20953 | CGCGCGCTCCA-----GCGCTATGGCGATCGCGCTC | 20982 | 1578 | LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis | 1597 |
| Qy | 1318 | alGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProG | 1338 | 22020 | CGTCTCTCGCTCTCGCTCGCTCGAGAAAGTCCCTCGCGCAGCATCGCCCGCGCGCGC | 22079 |
| Db | 20983 | TCCATCGGGCGGATCAACAGCCCTGTTCACGACGATCTCGGCGAGCCCTCGCGCTC | 21042 | 1598 | SerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGly | 1617 |
| Qy | 1338 | luArgHisSerProHis-----HisLeuLysGluGlnHisIleHisIleArg- | 1352 | 22080 | GGGACTCGCTTCTCGCTCACCTCGTCCA---AGCCCTCGCGGACATCGCCCTCGACGC | 22136 |
| Db | 21043 | GCGCGCTCTCGCGATCTGGAGTCCGAGGCGTCTTGGCGCTCAAGCTGAGTTACGAC | 21102 | 1618 | ValSerGlyValAspLeuTyr-----ArgSerHisIle | 1628 |
| Qy | 1353 | -----GlySerIle---ThrGlnGlyLeuProArgSerTyrValG | 1365 | 22137 | GCCCTT---GTGGCTCTTCAACCGCGCGCGCTTCTCGTGGACACATCCACCCCATCGC | 22193 |
| Db | 21103 | TTGCGCTCCCACTCGCGCAGGTTCGAGTCGATTCGCGAGAGCTCTCGATCTCTGTG | 21162 | 1629 | Pro-----LeuAlaPheAspProThrSerIleProArgGlyIlePro | 1642 |
| Qy | 1365 | luAlaGln-----GluAspTyrLeuArgA | 1373 | 22194 | CCATCCGAGCGAGCGGATGACCTGGGCGCTGGCGCGCTCGCTCGCGCTCGAGCA---CC | 22250 |
| Db | 21163 | TGGCTCGAGCGGCTCGAGCGGCTCCGTTCTACTCCACCGTGGAGCGCGCGCGATC | 21222 | 1643 | LeuAspAlaAlaAlaTyrTyrLeuProArgHis----- | 1654 |
| Qy | 1373 | rgGluAlaLysLeuLysArgGluGlyThrProProPro----- | 1387 | 22251 | CGAGCGCTGGGAGGGCTCGTTCGACGCTCGCGGACGATCGACCGCGCGCGCGCGCGG | 22310 |
| Db | 21223 | GACCGGAGCGAGCTCGACGCGCTACTGTGTACCGGAACCTTCGGCAGCGCGTTCG | 21282 | 1655 | LeuAlaProAsnProThr-----TyrProHisLeuTyrProPro---Tyr | 1668 |
| Qy | 1388 | -----ProProSerArg----- | 1391 | 22311 | CTTCTCGCGCTCTCGCGCTCGCGCAAGCATGAGGACCATCGCTCTCGCGCGCGCGG | 22370 |
| Db | 21283 | GCAGACGCTGTGCAAGCCCTCTTTCGCGGAGAACATCGCTTCTTCTGGAGGTGAGCCCC | 21342 | 1669 | LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp | 1688 |
| Qy | 1392 | -----AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeu----- | 1405 | 22371 | GTCTACCTCGCGCTCGCTCGCGCTTCCCTCGCGGACGCGCGCGCGC-----CGCACG | 22424 |
| Db | 21343 | AGTCTGTGTGCTGACCTTGGCTTGTGCACGAGCTCTCGAAGCGTTCGAGGCGCTCG | 21402 | 1689 | TyrIleThrSerGlnGln-----MethHisIleAsnThrAlaThrAlaMetAlaGlnArg | 1706 |
| Qy | 1406 | -----LysLeuLys-----ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA | 1423 | 22425 | TACCTTCAAGCCCGGAGGACCCCTCTCATCCCGGAG-GCACCGCGCGCGCTCGCGCTC | 22483 |
| Db | 21403 | GTGGTGGCTCTGTGTGGAGCGAGGAGGATCTACGGCGCTTCTCTGCTCTCGCTCTCC | 21462 | | | |


```
QY 505 lndInGlnGlnGlnInProMetPro-ArgSerSerGlnGlnGluLysAspGluLysGlu 524
Db 18478 AGCGCGCTCGCTCCCGCGCTCTCTCGCGATTTGCGCGAGCGCGCTCTCGAA 18537
QY 525 LysGluLysGluAlaGluLysGluGluGlu-LysProGluValGluAsnAspLysGluAs 544
Db 18538 GCCAGCGCGATCGCTCTCCGAGCAAGACGGGCGCACAGGCTCTCTCG 18586
QY 544 pLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaVa 564
Db 18586 ----- 18586
QY 564 lAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgLysThr-- 593
Db 18587 -----ACAAAGCTCCGAAACCGCTCGAGAGCGAGCAGATCCACCTG 18627
QY 584 -----ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaLe--ThrProGlnG 600
Db 18628 CTCTCTCGCTGTGTGCGCCACGAGCGGCTCTCTCTGCGGCATACCGAGCGCTCCAG 18687
QY 600 nSerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTrrThrGluGluGluMe 620
Db 18688 GTGACCCCC-----ACAAGGGCTTCATGACCTCGGCTCGAT 18726
QY 620 tGluThrAlaLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaAlaAlaArgMe 640
Db 18727 TCGCTCATGACCGTCGAGCTTCGTC-----GGCGCTTGACGAGCGCC 18768
QY 640 tValGlySerLysThrValSerGlnCysLysAsnPheTyr-PheAsnTyrLysLysArgG 660
Db 18769 ACCGGCATCAAGTCCCGGCCACCTCGCTTCGACCATCCCTCTCATCGCGTCGCG 18828
QY 660 lAsnLeuAspGluLysGlnGlnGlnHisLysLysLysMetGluLysGluArgAsnAla 680
Db 18829 CTCTTCTTGCGCA--CTCGCTCGCCAGCGCTCGGCGAGGCTCTCTCGTCGAGCGC 18885
QY 680 rArgLysLysLys-LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Db 18886 GAGCGCGCGCTCGCGCGCTCGCTCGCGAGCGAGCAG-----CCCATCGCC 18936
QY 700 valGluAspGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGlu 719
Db 18937 ATGCTCGGCATGCGCTCGCTCGCGGGCGCATCGCGCATGTCAGCGCTCTTTGGAG 18996
QY 720 -----GluAlaGluAlaLeu--HisAla-SerGlyAsnGluVa 731
Db 18997 TTCTCGCCCAAGGACGCGCGCTCGAGCCCATTCCTCCCATGCGGATGCGGTCGCGT 19056
QY 731 lProArg-----GlyGluCysSerGlyProAlaThrVa 742
Db 19057 GCCCTCTACGACCCGACCGCGAGCGCAAGGCCAAGAGTACGTCGCGCATGCGGCATG 19116
QY 742 lAsnAsnSerSerAspThrGluSerileProSer-----ProHi 755
Db 19117 CTGACGAGGTGACCTCTCTGATCTGCTGCTCTTTGGCATACGCTCGCGAGGCCAA 19176
QY 755 sThrGluAlaAlaLysAspThr--GlyGlnAsnGlyProLysProProAlaThrLeuG 774
Db 19177 TACTCGACCCCGACGACCGCTCTCTGATCTGCTGCTGCTGCTGCTGCTGAGCGCC 19236
QY 774 yAlaAspGlyProProProGly-----ProProThr----- 784
Db 19237 GGCATGCTCCCTCCACCTCAAGGATTCCTCCACCGGCTCTTCTGCGGCATCGCGCC 19296
QY 785 -----ProProArgArgThrSerArgAlaProileGluPr 796
Db 19297 AGCGAATACGACTGCGAAACACGAGCTCGAAGAGTGAAGCGTATGCTCCCAAGGC 19356
QY 796 oThrProAlaSerGluAlaThrGlyAla--ProThr----- 807
Db 19357 ACCGCGGGTCTTTCGCGGGGGGCTTGGCTTACAGCTCGCGCTGCAAGGGCGCG 19416
QY 808 -----ProProProAlaProProSerProSerAlaProProValProLy 824
```

| | | | |
|---|--|---------------|-------|
| PI | Lam ST, Hammer PE, Uknes SJ; | | |
| XX | WPI; 1996-040226/04. | | |
| DR | New genes for biosynthesis of anti-pathogenic substances - pref. | | |
| PT | pyrrolinitrin and soraphen, useful for disease control in plants. | | |
| PT | | | |
| XX | Claim 3; Page 140-158; 190pp; English. | | |
| PS | This is the soraphen gene cluster from S. cellulosum. The gene cluster | | |
| CC | may be expressed recombinantly to produce soraphen, or expressed in a | | |
| CC | transgenic plant for disease-resistance. (Updated on 16-OCT-2003 to | | |
| CC | standardise OS field) | | |
| XX | Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 U; 0 Other; | | |
| SQ | | | |
| Alignment Scores: | | | |
| Pred. No.: | 7.65e-11 | Length: | 28598 |
| Score: | 641.00 | Matches: | 661 |
| Percent Similarity: | 31.16% | Conservative: | 298 |
| Best Local Similarity: | 21.47% | Mismatches: | 1183 |
| Query Match: | 4.85% | Indels: | 941 |
| DB: | 2 | Gaps: | 140 |
| US-09-522-753-5 (1-2517) x AAT06769 (1-28598) | | | |
| Qy | 4 SerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProHisSer 23 | | |
| Db | 16988 TCGAGCGCTCAGCGCGCAGCTCCACGCGGACGCGGACGAGCGCGCGCTCGCC 17047 | | |
| Qy | 24 LeuSerTyrProValGlnIleAlaArgThrHisThr----- 35 | | |
| Db | 17048 TGCTCTTCCACCTCTCGAGCTTCCACACGCGCCAGACGACGAGCGTGCACA 17107 | | |
| Qy | 36 -----AspValGlyLeuLeu-GlutYrGlnHisHisSerArgAspTyrAl 50 | | |
| Db | 17108 CTGGCGCTACCGATACGTGGAGGCTCTGACACCGCGCCACGCGCGACCTCG 17167 | | |
| Qy | 50 aSerHisLeuSerProGlySerIlelleGlnProGlnArgArg----- 65 | | |
| Db | 17168 CCGGACCTGCTCTCTGCGCTCGCGCTCGCGCTCGCGCGCGCGCTCGCTCC 17227 | | |
| Qy | 66 -----ProSerLeuLeuSerGluPheGlnPr 74 | | |
| Db | 17228 TCACGATGCGCTTACCGCGCGCGCGCGCTGCTCGCTCGCGCTGAG-----CC 17281 | | |
| Qy | 74 oGlyAenGluArgSerGlnGluLeuHisLeuArg----- 85 | | |
| Db | 17282 AGGTTACATAGCGCGCGGCTCTCACGAGCACCTCGCGAGGCTTTGCCGAGACTG 17341 | | |
| Qy | 86 -ProGluSerHisSerTyrLeuProGluLeuGlySerGluMetGluPheleGluSe 105 | | |
| Db | 17342 CCGCGATTCGCGCGCTGCTCTCCCT----- 17366 | | |
| Qy | 105 rlyAsnProArgLeu-----GluLeuLeuProAspPr 116 | | |
| Db | 17367 -CTCGCCCTCAGCAGCGCCCTCGCGACCATCGCGCTCGCGCGCGCTTGCC 17425 | | |
| Qy | 116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer---GluAspLe 135 | | |
| Db | 17426 TCTCGCTCGCTCGTCCAGCGCTCGCGACCTCGCGCTCGCGCTCGCTGTCCT 17485 | | |
| Qy | 135 uThrlyAsnArgSerLeuThrGlyLeuLeuGluProValSerProSerProProHi 155 | | |
| Db | 17486 TCACGCGCGCGCGCTCTCGATTGAGACACTCCAGACCTCCCATCCACCGAGCCA 17545 | | |
| Qy | 155 sThrAspProGluLeuGluValProProArgLeuSerlySerGluLeuLeuGlnAs 175 | | |
| Db | 17546 T-----GATCGGCTTGGCGCGCTCGTCCGCTCGA----- 17579 | | |
| Qy | 175 mMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLy 195 | | |
| Db | 17580 -----GCACCCCGAGCGGTGGCGCGGCTCGTGC 17608 | | |
| Qy | 195 slyGlnGlnGlnLeuGluGluAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 215 | | |
| Db | 17609 ACCTCGCGCAGCGCTCGACGCGGCGCGCGCGCTTCTCGCGCCCTCGCCACG 17668 | | |
| Qy | 215 rProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235 | | |
| Db | 17669 GCCA-----CGACG 17677 | | |
| Qy | 235 glySylAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255 | | |
| Db | 17678 AAGACAGCTCGCTCGCGCC-----GCCCGGCTCTACGACGCC 17719 | | |
| Qy | 255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275 | | |
| Db | 17720 GTTCTCGCGCGCGCTCGG-----CGATCGGC 17749 | | |
| Qy | 275 aMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrly 295 | | |
| Db | 17750 CTGCGCTCGCGCTT-----CATGCCCGAGGACCATCTCATCCGCTGGTA 17800 | | |
| Qy | 295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLys-----Ly 312 | | |
| Db | 17801 CCGCGCCATTCGCGCTCACGTCGCCGATGCTCGTCAAAAGCGCTGACACCTCG 17860 | | |
| Qy | 312 sValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysValArgGluTy 332 | | |
| Db | 17861 TCTCATACGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 17920 | | |
| Qy | 332 rTyrGluLysGlnPheProGluIleArgLysGlnArg---GluLeuGlnGluArgMetGl 351 | | |
| Db | 17921 CGCCCTCGCGCGCGCTCACCTTCGCGCGCTGCGATGTCGCCGACGAGCGCTGTG 17980 | | |
| Qy | 351 n-----SerArgValGlyGlnArg-GlySerGlyLeuSerMetSer----- 364 | | |
| Db | 17981 CCAGCTTCTCGAGAGCTCGACCGCGA---GGCCACAGGTGAGCGCGCTGTCCACG 18037 | | |
| Qy | 365 -----AlaAlaArgSerGluHisGluValSerGluI 375 | | |
| Db | 18038 CGGCGCATCGAGCCCGCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18097 | | |
| Qy | 375 leIleAspGlyLeuSerGluGlnGlnAsnLeuLysGlnMetArgGlnLeuAlaVal 395 | | |
| Db | 18098 TTGCTCTCGCGAAGTACAGGT-----GCAAGACACCTCCACGACCTCGCTCG 18148 | | |
| Qy | 395 leProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyL 415 | | |
| Db | 18149 CTCGACCCCTC----- 18159 | | |
| Qy | 415 euMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluG 435 | | |
| Db | 18160 -----GACGCTTTGTTCTCTCTCGCTCGCGCGCGGTC-----GTCTGGGCGCG 18205 | | |
| Qy | 435 lnGluLys-----GluThrPheArgGluLysPheMetGlnHisProL 449 | | |
| Db | 18206 GACAAAGAGCGGTATGCGCTGCGCTGCGACCGCTTCTCGATGCTCGCGCGCGCGCG 18265 | | |
| Qy | 449 ysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuT 469 | | |
| Db | 18266 GCAGCTTGGCTGACGCGCGACA-----TCGCTGGCCT 18298 | | |
| Qy | 469 yTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer----- 481 | | |
| Db | 18299 GGGG-CGTGTGGCGCGCGCGGATCGCTACCGGCTCTTGGCAGCCCGAGTAGACAA 18357 | | |
| Qy | 482 -----LeuValArgArgSerTyrArgArgGlyLysSerG 494 | | |
| Db | 18358 CGCGCTGTGCGCGATGGCCCTCTCGTGGCGCTGCGCGCTCGCGCTCGCGCTCGAG 18417 | | |
| Qy | 494 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 505 | | |
| Db | 18418 CACGACGAGACCCCTCACCGTCGCCGACATCGACTGGGCGCGCTTGTGGCTTCGTT 18477 | | |

QY 1448 -----GluGlySerIleThrGlnGlyThrProLeu 1457
Db 13035 GCAACTCTCTCCCGAGTCCGCCACACAGCGGAGAGCACTCTCCACCGCGGTCTGAAT 13094
QY 1457 uLyTyAspThrGlyAlaSerThrGlySerLysLysHisAspValArgSerLeu11 1477
Db 13095 CCGGACCAAC-----GCCGTGAGTCCGGCCGCCATCCGGGCCCGGTGGCGCCCTGAC 13148
QY 1477 eGlySer-----ProGlyArgThrPheProProValHisProLeuAspVa 1492
Db 13149 CGGGCA-GCAGCAGGGCCACACCGGGCGCCCTCTCTCTCCGTACCGCCCT----- 13199
QY 1492 lMetAlaAspAlaArgAlaLeuGluAlaCysTyThrGluGluSer-----LeuLy 1509
Db 13200 -----CGTCCGCATCTTCCGCCCTCTCTCTCCGCCCGACTCTCTCCGGTACCGA 13249
QY 1509 sSerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaProVal11 1529
Db 13250 GCCTCGCCCTTGGCTCGACGGCGGTGACTCCCGCTCCAGCGCGGCCCGCGGGC 13309
QY 1529 eValProGluLeuGlyLysProArgGlnSerProLeuThrTyThrGluAspHisGlyAlaPr 1549
Db 13310 GCCACCGAGCTC---AAGCCCGCTTCAGGGACGTC---TTCAGAGTCCGGTTCGAGGC 13363
QY 1549 oPheAlaGlyHisLeuProArgGlySerProVal-----ThrMetArg1 1564
Db 13364 CTGAACGGACCGCGCTCCAGGAGCCCAAGGGGGTCTGTCGAGTGAACCCAGCAAGTC 13423
QY 1564 uProThrProArg-----LeuGlnGlySerLeuSerSer----- 1576
Db 13424 GCCGTACCGGCATCGCGTGGTGGCTCCGGAGGATGGCGCAAGCGCTACTGGAG 13483
QY 1577 -----SerLysAlaSerGlnAspArg----- 1583
Db 13484 CAGTCACTCCGAGCACCGCCAGCGCGCCATCTCTTTCGACGCTCGCCCTC 13543
QY 1584 -----LysLeuThrSerThrProArgGluLeuAla----- 1593
Db 13544 CGCTCGCGATCGCGCGAGTGCATTCACCCCGCGGGCGGCGCTGAGCCCGCGC 13603
QY 1594 -----LysSerProHisSerThrValProGluHisHisProHisPro11 1608
Db 13604 GAGTCCGCGCATGGACCGCGCGCGCAGTTCG---CCGTGTCGAGCGCGAGGAGC 13660
QY 1608 eSerProTyThrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyArgSerHis11 1628
Db 13661 CTCGCCGACAGCGAC----- 13676
QY 1628 eProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAl 1648
Db 13677 -----TCGACGTCGCGACCTCGACCCCGGATCGGGTGGAGCATCGGAGCGCC 13729
QY 1648 aTyTyLeuProArgHisLeuAlaProAsnProThrTyProHisLeuTyProProTy 1668
Db 13730 GTGGCGGGACCA-----CTCGCTG 13750
QY 1668 rLeuIleArgGlyTyProAsp---ThrAlaAlaLeuGluAsnArgGlnThrIleAs 1687
Db 13751 GAACGCGAGTACCTCGCCCTCAGCGACAGCGGGCGACAGTGGGAACCTCGACCTCTCTAC 13810
QY 1687 nAspTyIleThrSer-GlnGlnMethHisHisAsnThrAlaThrAlaMetAlaGlnArg 1707
Db 13811 CTCTCGCGGACCTCTAGAGCGCTTACCCCGAGTCGCTCGCCCGGAGGTGGCGGG 13870
QY 1707 laAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyThrAlaAlaG 1727
Db 13871 GTGATCGG---CGCGGAGG-GGCGCGGCGGAGTCTCT----- 13904
QY 1727 lyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1747
Db 13905 -----CCACCGGTGCACCTCCGG 13923

1747 hr-----ProGlyThrProAlaThrAlaMetAspArgLeuAlaTyLeuProThr- 1763
13924 CATCGACTCGTCGGCCACGCGCGACCTCATCGCGAGGCGAGCG-----CCGACGT 13977
QY 1764 -----AlaProGlnProPheSerSerArgHisSerSerSerProLeuSerP 1779
Db 13978 CGTCTCGGGCGGGCGGACGACACCC---CCATCTCGCATCGCGTCTCGCTTCGA 14034
QY 1779 roGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgA 1799
Db 14035 CG-----CATCAAGCCACCTCGCCCGACCAACAGACCGCG-GCGCAGCCTCGCGCC 14087
QY 1799 spArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThr 1819
Db 14088 CTTTCGACAGGAAACGAA-CGGCTTCGTCTCGC----- 14121
QY 1819 hrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerG 1839
Db 14122 -----GGAGGGCGCGCGCC 14134
QY 1839 lySerSerGlyGlyGlyGlySerSerSerArgPro----- 1851
Db 14135 GTCTCTGTCTGAGGAGCTGGGGCGACGCGGGCGCGCGACGCCACGCTCTACGCTCTC 14194
QY 1852 -----AlaSerHisSerHisAlaHis----- 1858
Db 14195 GTCTCCGGCTACGCCACGCGCTGCAACGGCTACACATGACCGGGCTCACCCGCGACGGC 14254
QY 1859 -----GlnHisSerProIleSerProArgThrGln-----A 1869
Db 14255 AGGAGATGGCCGAGGCGATCCGGCGACGCTCGCGAGCGGGACCGACCGCGCGCC 14314
QY 1869 spAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrA 1889
Db 14315 GTGGACTTACTCAACGCCCGCTCGGCACCAA-----G 14350
QY 1889 laValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgP 1909
Db 14351 CAGAACGACCGGCACGAGACCGCGCTTCAAGCCACGCTCGCGAGCGGGCGAGGAGC 14410
QY 1909 roAlaAlaThrPheProProAlaThrHisCysProLeuGly----- 1922
Db 14411 GTCCGCTCAGTCCATCAAGTCAGTCGGCACCTCGTCTGGTGCATCGGTTCGCTG 14470
QY 1923 -----GlyThrLeuAspGlyVal-----TyrProT 1931
Db 14471 GAGATCCGCGCTCGGGCGCTCGCATGAGTACGGCGTGTGGCGCCACCGCAACCTC 14530
QY 1931 hrLeuMetGluProValLeuLeuProLys---GluAlaProArgValAlaArg----- 1947
Db 14531 GACACGCGCGACCGCGAGTCCGACTGACTGTCGCCCGACGAGGCGCGAGACGCGG 14590
QY 1948 -----ProGluArg-----ProA 1952
Db 14591 GTGGACAGGTCTCGAGCGTCCGCGAGCGGTTCGCGGTTTCAGAGCGCATGGTCTC 14650
QY 1952 rg-AlaAspThrGly-----HisAlaPheLeuAlaLysPro--- 1963
Db 14651 ACCCGGACACGGGGCGCCCTCCCGACGGCTGACACCGCTCACCGCAGCGCCCGCA 14710
QY 1964 ProAlaArg---SerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
Db 14711 CCGCGCGCGAGCGCGCGGAGTCCCGACCGCTCGCACGGAAGGACCCCGCCATGA 14770
QY 1983 ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeu 2002
Db 14771 CCA-----CGGCC-----CCTCCGCGACCGCCCGCAGG 14797
QY 2003 AlaProHisHisAlaSerProAspProAlaProAlaSerAlaSerAspProHis 2022
Db 14798 GCGCCCTCCCGGGCGCGCTTCCCGCTTTCACCGCATCGGCTCGCGCCCGC--- 14854
QY 2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuArgSerLeuGly 2042

Db 10964 ATCCAGCGCTGCGGAGC-|||:|:|AGGATCGCGCGGAAC 10999
QY 937 uLeu---SerProArgProSer--LeuLeuThrProThrGlyAspPro----- 951
Db 11000 GTGTCACCGCCCTGTTCATCCGAGCGTCCGGCGTACCGGCGCGGCACACCGCGCC 11059
QY 952 -----ArgAlaAsnAlaSerProGlnLysProLeuAspLeu-----Ly 964
Db 11060 GCGCGGAGATCCCGCGCGCGCGGAAGGCCACACACAGCGCTCCCGCGCGCTCGCG 11119
QY 964 sGlnLeuLysGlnArg-----AlaAlaAlaI 973
Db 11120 AAGACACGAGAGCGGTGCGGTCCGACGCTCCACGAGATCCGTATCCGCTCCGTGCC 11179
QY 973 eProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThr-- 992
Db 11180 ACCACAGAGCGTCCGGGGCGTCCCGCGCGCGCACGCGTCCGAGGAGCGGCAGCGG 11239
QY 993 -----LysProAlaProProAlaProProAlaProProGlnAsnLeuG1 1006
Db 11240 TGGCAGAAACCGCGCGAGGCGCGGACACGCTCAGCGCCGCGCGCGCGCGAGG 11299
QY 1006 nProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerPr 1026
Db 11300 TCCTCGCGGAGGACGCGCGAGCGGGCGGTCTGCTCCAGCGCGCACATGCTCGCCACC 11359
QY 1026 oAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPr 1046
Db 11360 AGCACGATCATCGCTCGCGCA--GCACCCCGCGCGCGAGCGCTCTCGGCC 11416
QY 1046 oPro----- 1047
Db 11417 GCCCGCGCCCATGAACCAAGGTCTCGTCGGCGCGCGGAAGCGCGCGCACGATA 11476
QY 1048 ----CysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSe 1066
Db 11477 CGCTCGGGTCTCTATCCACTCCGGGTGTCAGCGACCGTGCGCACACAGTTCTGGCGTTG 11536
QY 1066 rProHis-AlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProL 1086
Db 11537 TCCACACTGCGCGCGCGCGGTACC-----CCCCACCCCAAGAGC 11578
QY 1086 euGlyLeuHisAspThrAlaArg-----ProVal 1096
Db 11579 CTGCTGTGCACGACAGCGCGCTACCGAGCATCTCCGCGCTCGCGGTCACTGTCCCTCCC 11638
QY 1096 euPro-----ArgProProThrIleSerAsnProProProLeuIleS 1110
Db 11639 TCCCGCCAGGGCGCGTGCCTCCGCGCGCTGACAGCCTCCGGCGCGCTTTGAGCCCG 11698
QY 1110 erSerAlaLysHis-----ProSerValLeuGluArgGlnI 1122
Db 11699 GATCGAGCCGAGATCGAGACCCCGCGGATGTGGGGCGTCACTCCC--CGSGCGCGG 11757
QY 1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHis---- 1140
Db 11758 CGGGGCGCGAGGGGACCGGG-----CTCGGGTCCCAGGAAGAACACACGC 11805
QY 1140 ----- 1140
Db 11806 GGCATGAGCGCCTTCACTCTCACGAGTTCAAGAGCTGTCGAGCAGAGCTACGACGCC 11865
QY 1141 ----AlaLysAlaPro-----ValGlyPro-----ValThrMetGlyLeuP 1153
Db 11866 GAATCGCCGAGCCCTCCACGGGCGAGGCCCTCGACACGAGCTTACCAGTCTGGGTAC 11925
QY 1153 roLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuS 1173
Db 11926 GACTCGTGAGCGGTCTACGAGATCGTCACCGCATCCAGGACGAGCAGCGAGTCAACCGTG 11985
QY 1173 erProArgGlyGlnAlaGlyProPro-----G 1182
Db 11975 GCGCGACAGGGCCCGCTTCCAGCGCGCTCTCCGCGCGCGCGCTCCAGCAGCACCGCGCGC 13034

11986 CCGCAGCAGGA---GCTGGACCTCTCTGACACCCCGCGCGCCCTCATCGGTACGTCGAC 13042
QY 1182 luSerLeuGlyValPro-ThrAlaGlnGlu----- 1191
Db 12043 GCGCGCGCGGTTCGCGACCTGACCCGACGGAGGGTGGGGAGGGGGCGGCGGAGC 12102
QY 1192 -----AlaSerValLeuArgGlyThrAlaLeuGlySer----- 1202
Db 12103 GGGCGGGGCGCGCGCGGTACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAGCGCCC 12162
QY 1203 ValPro-----GlyGlySerIleThrLys 1210
Db 12163 GTCCCTAGACCTGACCGCGCTGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTCCGC 12222
QY 1211 GlyIleProSerThrArg-----ValProSerAspSerAla----- 1222
Db 12223 GGGTCCCGGAGCGCGCGGAGGAGCGACACACGCGCGCGCGCGCGCGCGCGCGCGC 12282
QY 1223 -----IleThrTyrArgGlySerIleThr-----His 1231
Db 12283 CGGATGACGCGCGCGCGCGCGCGAGACTGCCGTCGCGCGAGCGCTCGAGCAGCACAC 12342
QY 1232 GlyThrProAlaaspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAaspSer 1251
Db 12343 GCCACGCCCGCC-----GCCCGTGT----- 12363
QY 1252 ProSerArgLeuAspArgGlyArgGluAaspSerLeuProLysGlyHisValIleTyrGlu 1271
Db 12364 ---CGTCGCGCTCGCGCTCGCGAGCAGGTTCGCCAGCGCGCGCGCGCGCGAGC 12420
QY 1272 GlyLysGlyHisValLeuSerTyrGluGlyGlyMet----- 1284
Db 12421 GGG---GGCGGCGAGTGCACCGACAGAACCGCGCGCGCACCGCTCGCGCGCGGTGAG 12477
QY 1285 SerValThrGlnCysSerLysGluAaspGlyArgSerSerGlyProProHisGluThr 1304
Db 12478 CCGCGCACCGCTCG 12537
QY 1305 -----AlaAlaProLysArgThrTyrAsp-----MetMetGluGly 1316
Db 12538 TTCGCGCTCGCCCTCGTCAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 12597
QY 1317 ArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIlePro 1336
Db 12598 TAGTCT---CGCTGAGCGCGCGCGCGCGCGTCAAGCCGCTCGTGGCGCGAGCG-- 12645
QY 1337 ProGluArgHiserProHisLeuLysGluGlnHisIleArgGlySerIleThr 1356
Db 12646 CCGCGCACCGCGCGCGCGCGCGCGCTCCGCGCGCGCACAGCAC-----CGTCTGAGC 12696
QY 1357 GlnGlyIleProArgSerTyr-ValGluAlaGlnGluAaspTyrLeuArgArgGluAlaLy 1376
Db 12697 CGGCGCGTTCACCG 12756
QY 1376 sLeuLeuLysArgGluGlyThrProProProProProProProProProProProPro 1396
Db 12757 CCGCTC 12816
QY 1396 aTyrLysThrGlnAlaLeuGlyProLeuLysLysProAlaHisGluGlyLeuValAl 1416
Db 12817 GCGGACG 12867
QY 1416 aThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisTh 1436
Db 12868 -----GGCGCGGAGTTCCCGCAGCGTGTGCCCGAGCAACAGGTGCGCGCGCAC 12914
QY 1436 r-----ProGlu-----LeuPro----- 1440
Db 12915 CTTGCT 12974
QY 1441 -----LeuAlaProArgProLeuLys----- 1447
Db 12975 GCGCGACAGGGCCCGCTTCCAGCGCGCTCTCCGCGCGCGCGCGCGCGCGCGCGCGC 13034


```
QY 2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256
Db 15456 CGCTACTCCCTTCGACCGCGTGCCTCGAC-----CGTCGT 15497
QY 2257 MetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThr 2276
Db 15498 CGCGAGGCGCGCCTCTCACCCTGGAGACACCCGCGACCGGAGGAGCGCAGCG 15557
QY 2277 GluSerAsnSerAlaMetValLysSerLysGlnGluLeuAsn-----LysLysLeu 2294
Db 15558 ACCCGGATACCGGGAACCTCGCGGTACCGCGGCACCTTCGACCCCGCGCGCTC 15617
QY 2295 AsnThr-----HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
Db 15618 CGACGCGCCCGCGGCTCGAACGCGCGCGCTCGCCCTCGCGACCGCGGCGCTGCG 15677
QY 2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArg----- 2328
Db 15678 ACCCG---GGAGCTGACGCTCTTCGCGGAGCGCGCGGGCTCCCGCGCGCAGCG 15734
QY 2329 -----SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeu 2342
Db 15735 CGCGAGCGCGCGCTCGCGCGCTCTTCGCGCGCGCGCGGTTCGCGTGAGCGTGCC 15794
QY 2343 GluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProPro 2362
Db 15795 GAAGACCAACCGCGCGCTCGCTCGCGCGCGCGCGCGCGCTCG-AGTCGCGCGCGCGCC 15853
QY 2363 LeuSer-----AlaAsnAlaPheAsnProLeuAsnAlaSerLeuPro 2378
Db 15854 TCCTGCGCTCGCGCGCGCTCGTGCCTCGCGCGCGTCCACCTCGACGAGTCCGTC 15913
QY 2379 -----AlaAlaMetPro----- 2382
Db 15914 CGTACGCGCTCGACCTCGTCCGCGACACCCCGCGCGCTCCCGCTCGCACGCGCTCG 15973
QY 2383 ---IleThrAlaAlaPheGlyArgSerAspHisThrLeuThrSerProGlyGlyGly 2401
Db 15974 TCCTGCGCGCGCGCGCGGCTTCAACGCGCGCGCTCGTCTCGCGCGCGCGCGCG 16033
QY 2402 LysAlaLysValSerGlyArgProSerSerArg-----LysAlaLysSerProAla 2418
Db 16034 CCGGTACCGCTGAGGCGCTCGCCCGGCTCGCGCGCGCGCGCGCGCGCGTCCACCG 16093
QY 2419 ProGlyLeuAlaSerGlyAspArgPro----- 2427
Db 16094 CCGCGCACACCCACAGAGTTCCTATGACGCGACTGACCGTGCAGCAGCTCAAGAAC 16153
QY 2427 ----- 2427
Db 16154 TGCTCGCGGAGACCGCGCGGAGGACACAGCTCGACCTCGCGGAGAACTCGACACGC 16213
QY 2428 ProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrPro----- 2444
Db 16214 CCTCTGGGACCTCGGCTACGACTCCCTCGCGCTGTGTGAGACGCGCGCGTGTCCAGC 16273
QY 2445 -----LeuThrAsnArgVal-----TrpGluAspArgProSer 2455
Db 16274 AGCGGTACGCATCGCGCTGACTGACGAGACGCTCGCGCGGTGGGACCCCGCGCGAAC 16333
QY 2456 ---SerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGly 2474
Db 16334 TGCTCGAGGAGTCAACACCA---CCCGCGCACCGCTGAGCGCGCGCGCGCGAAC 16390
QY 2475 ValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu-----Ala 2492
Db 16391 ---GAAGGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16447
QY 2493 GlyProHis-----HisAlaTyr 2498
Db 16448 GGACCTCACAGCAAGGTGCGGTCTCTGG 16474
```

RESULT 52

```
AAV25925
ID AAV25925 standard; cDNA; 24379 BP.
XX AC AAV25925;
XX 15-JUL-1998 (first entry)
XX Streptomyces roseofulvus frenolicin gene cluster.
XX Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B;
XX antibiotic; ss.
XX Streptomyces roseofulvus.
XX Key Location/Qualifiers
FH CDS 636..2948
FT /tag= a
FT /transl_except= (pos:636..638,aa:Met)
FT /note= "encodes protein given in AAW55800"
FT 2945..3916
FT /tag= b
FT /transl_except= (pos:2945..2947,aa:Met)
FT 4020..4844
FT /tag= c
FT /transl_except= (pos:4020..4022,aa:Met)
FT /note= "encodes protein given in AAW55802"
FT 4841..6415
FT /tag= d
FT /note= "encodes protein given in AAW55803"
FT 6533..7183
FT /tag= e
FT /transl_except= (pos:6533..6535,aa:Met)
FT /note= "encodes protein given in AAW55804"
FT 7344..8897
FT /tag= f
FT /note= "encodes protein given in AAW55805"
FT 9164..10012
FT /tag= g
FT /transl_except= (pos:9164..9166,aa:Met)
FT /note= "encodes protein given in AAW55806"
FT complement(10105..10621)
FT /tag= h
FT /transl_except= (pos:10621..10619,aa:Met)
FT /note= "encodes protein given in AAW55807"
FT complement(10618..11628)
FT /tag= i
FT /transl_except= (pos:11628..11626,aa:Met)
FT 11809..12066
FT /tag= j
FT /note= "encodes protein given in AAW55809"
FT complement(12154..13209)
FT /tag= k
FT /transl_except= (pos:13209..13207,aa:Met)
FT /note= "encodes protein given in AAW55810"
FT 13409..14686
FT /tag= l
FT /transl_except= (pos:13409..13411,aa:Met)
FT /note= "encodes protein given in AAW55811"
FT 14767..16047
FT /tag= m
FT /note= "encodes protein given in AAW55812"
FT 16120..16371
FT /tag= n
FT /note= "encodes protein given in AAW55813"
FT complement(16453..16935)
FT /tag= o
FT /transl_except= (pos:16935..16933,aa:Met)
FT /note= "encodes protein given in AAW55814"
FT 17088..18903
FT /tag= p
```



```

13661 CTGCGCGACAGCGGAC----- 13676
1628 eProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAl 1648
13677 -----TGCAGCTCGCGGACCTCGAGCCCGACCGATCGGGGTGAGCATCGGCGCGCC 13729
1648 atyTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProPro 1668
13730 GTGGCGGGACCA-----CCTCGCTG 13750
1668 rleuIleArgGlyTyrProAsp---ThrAlaAlaLeuGluAsnArgGlnThrIleIleAs 1687
13751 GAACGCGAGTACTCGCCCTCAGCAGCAGCGGCGCAGTGGGAACCTCGACCTCTCTATC 13810
1687 nasPTyIleThrSer-GlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgA 1707
13811 CTCTCGCGGACCTCTACGAGCGCTTCACCCCGAGCTCGCTCCCGCGAGGTGGCGGG 13870
1707 laAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaG 1727
13871 GTGATCGG--CGCGGAGG-GGCGCGGCGGCGAGTCT----- 13904
1727 lyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1747
13905 -----CCACCGGCTGCACCTCCCG 13923
1747 hr-----ProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThr- 1763
13924 CATCGACTCGCTCGGCGACCGCGGACCTCATCGCGGAGGCGAGC-----CCGACGT 13977
1764 -----AlaProGlnProPheSerSerArgHisSerSerSerProLeuSerP 1779
13978 CGTGCTCGCGGGCGGACCGACCC---CCATCTCGCGATCGCGCTCGCTTCTGA 14034
1779 roGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgA 1799
14035 CG-----CCATCAAGCGCCACTCGCGCCCAACAGCAGCAGCCG-GGCGACGCTCGCGCC 14087
1799 spArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThr 1819
14088 CCTTCACAGGGAACGAA-CGGCTTCGTCTCGC----- 14121
1819 hrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerg 1839
14122 -----GGAGGCGCGCGCC 14134
1839 lySerSerGlyGlyGlyGlySerSerSerArgPro----- 1851
14135 GTCTCTGTGTGAGGAGCTGGGGCACGCGCGGCGCGCGACGCCACACGTCTACGCTCTC 14194
1852 -----AlaSerHisSerHisAlaHis----- 1858
14195 GTCTCGGCTACCGCAGCGCTGCACCGGTACACATGACCGGGCTCACCCGCGCAGCGC 14254
1859 -----GlnHisSerProIleSerProArgThrGln-----A 1869
14255 AGGAGATGCGCGAGCGATCGCGCAGCGCTCGCGGAGCGGACCGACCCCGCGCGCC 14314
1869 spAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrA 1889
14315 GTGGACTTACGTCAACGCCCAACGCGCTCGGCACCAA-----G 14350
1889 laValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgP 1909
14351 CAGAACCGCGGACGAGACCGCGCTTCAGGCCACGCTCGCGGAGCGGCGGCGAGC 14410
1909 roAlaAlaThrPheProAlaThrHisCysProLeuGly----- 1922
14411 GTCGCGGTACGCTCCATCAGTCATGATCGATCGGCGCACTCGCTCGGTCCATCGGTTCGCTG 14470
1923 -----GlyThrLeuAspGlyVal-----TyPro 1931
```

```

14471 GAGATCGCGCGCTGGCGGCTCGCCATGAGTAGCGGCTGGTCCGCCCCACCGCAAACTC 14530
1931 hrLeuMetGluProValLeuLeuProLys---GluAlaProArgValAlaArg----- 1947
14531 GACACGCCCGACCCCGAGTGGGACCTCGACTAGTGCCTCCCGACGAGCGCGGAGACGG 14590
1948 -----ProGluArg-----ProA 1952
14591 GTGGACAGGCTCTGAGCGTTCGGCAGCGGCTTCGGCGGTTCAGAGCGCATGCTCTC 14650
1952 rg-AlaAspThrGly-----HisAlaPheLeuAlaLysPro--- 1963
14651 ACCCGGACAGCGGCGCCCGCTCCCGACGCGCTGACCCCGCTACCGCAGCCCGCGCA 14710
1964 ProAlaArg---SerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
14711 CCGGCCGCGAGCGCGCGGAGTCCCGCAGCTCCGACACGAGGACCGCCCATGA 14770
1983 ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeu 2002
14771 CCA-----CCGCGC-----CCTCCCGCAGCGCCAGG 14797
2003 AlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHis 2022
14798 GCGCCCTCCCGGCGCGCTTCGCGCGCTTCACCGCATCGGCTCGCGCC--- 14854
2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly 2042
14855 -----CCACCGGCTCGGACCGAGGAGTGG-TGGGCGGCCACCTCGCGC-----GGG 14901
2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerPro 2062
14902 GAGCAC-----GGTCTCGCGCGCTCACGAGTAGCAGCGGAGC 14940
2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuLeuAspLysSer---His 2081
14941 GGC-----CACCCCGGTGGACTCTCGCGCGCTCCCGGACTTCAGACCGCGCGCCAC 14994
2082 LeuGluGlyGluLeuArgProLys----- 2089
14995 CTGCGCGCGCGCTCTCTCGCGCA-GACCACCGGCTCACCGGCTCGGCTCGTCCGCGC 15053
2090 -----GlnProGlyPro---ValLysLeuGlyGlyGlu 2099
15054 CGACGAGGCACTGAGGAGCGCGCGCTCGACCGCGCGCGCTCGCGAGTACGCGGCGAG 15113
2100 AlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeu 2119
15114 CGCGGTCACTCCACGCCACCGCGCGCTTCGAGTTCCACCCCGGAGATCCGCAAACT 15173
2120 LeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAla----- 2136
15174 CTGGACGAGGCGCGCGCGCTCAGCGTCTACGAGTCTTCGCTCGCTGTTCTACGCGCT 15233
2137 GlnHisLeuSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSer 2156
15234 CAACACCGCGCAGATCTCATCCGCGCAGCATCGCGCGCGCGCGCGCGCTCGTCTGTCG 15293
2157 AlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeu 2176
15294 CGACCGCGCGCGCGCT----- 15311
2177 ArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGly 2196
15312 CGACGCGCT-----CGGCCAGGCGCGCGCTACTGCGCAAGGCGGCGGT 15356
2197 SerProHisSerGluGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216
15357 GCT-----GGCGGTGAGCGCGG---CGTGGAGTCCGCGCTCGACCCCTCGGG 15401
2217 GlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis 2236
15402 CTGGCGCGCGCCACGCTC-----GTGCGGACCCCTCAGCGCTCCGCGGACCGCGCCAC 15455
```

[illegible]

| | | | |
|----|------|--|------|
| QY | 14 | AlaThrGluProArgTyrProHisSerLeuSerTyrProValGlnIleAlaArgThr | 33 |
| DB | 7955 | CGCGACGCGGGGAGGACCGCGCATCGA-----CGTCGCGGGCGCCCT | 7999 |
| QY | 34 | HisThrAspVal-----GlyLeu | 39 |
| DB | 8000 | CACCGTAGCGCTCGGTCTCTCGCGCTGCTCTCGCATCGTCAGAGCGCCGACGG | 8059 |
| QY | 40 | LeuGluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIle | 59 |
| DB | 8060 | CTGGGGCTCGCGCGCACCTCGTCCC----- | 8086 |
| QY | 60 | GlnProGlnArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSer | 79 |
| DB | 8087 | GCTCTCGGCGCCTCGCCCTGCTCGTTCGTCTCGTCAGAGCGGGTTCGCGCC | 8146 |
| QY | 80 | GlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGlu | 99 |
| DB | 8147 | GCAGCCGCTCAT-----ACCGCTCGGCATCTTCGGTCCCGCTCGTCGGCGGCAA | 8200 |
| QY | 100 | MetGluPheIleGluSerLysArgProArgLeuGlu-----LeuLeuProAspProLeu | 117 |
| DB | 8201 | CGTCGTGGCGATGACACGCGCGCGGCTCTTCAGCATGTCTTACTTCTCACCCCTCT | 8260 |
| QY | 118 | -----LeuArgProSerProLeuLeuAlaThr | 126 |
| DB | 8261 | CCTCAACACGAGTACGGGACTACAGCCCGCTCGCACCGGCT-TCGCTATCTCGCGGCTCG | 8319 |
| QY | 127 | GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu | 146 |
| DB | 8320 | CGTCGCCATCATGTCGCCCGCGAGTCTCCGCGCGCTCTGCGCGTCTCG- -GCC | 8376 |
| QY | 147 | ProValSerProProSerProHisThrAspProGluLeuGluLeuValProProArg | 166 |
| DB | 8377 | CCCGTACGACCCTGCTGTGTCATGGCTCTCACCGCGCGGACTGTCTGGTCTCTCGC | 8436 |
| QY | 167 | LeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetVal | 186 |
| DB | 8437 | GGCTCACCGAGACTCCGGGTTCGGGGGGGACTCTCGGCCCGGACCTGTGTCTCGCA | 8496 |
| QY | 187 | GluGlnGlnIleSerLysLysLysLysGlnGlnGlnLeuGluGluAlaAlaLys | 206 |
| DB | 8497 | TCGGACAGGCATCTCGATGTCGGCTCCGGCATCGCGGGGTTCGCGGGTTCGCGCGC | 8555 |
| QY | 207 | ProProGluProGluLysProValSerProPro----- | 217 |
| DB | 8557 | -----AGCAGGCGCGGCTCGCTCCGGGCTGCTCAACGCGACCGCGCAGCTCG | 8604 |
| QY | 218 | -----ProIleGluSerLys-HisArgSerLeuValGlnIleIleTyrAspGluAsnAr | 235 |
| DB | 8605 | GCGGCGCCTTCGGCTTCGGGTGTCGCGCGCTCGCCACTCGCGCGCGCGGACTCC | 8666 |
| QY | 235 | glysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr | 255 |
| DB | 8665 | TCGACGGGTTCGCGGGGCCACGGCGGAATCGCCCGGACGACGCCAGGCGTCCGGGACC | 8722 |

QY 136 ThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHis 155
 |||||
 Db 305 ACCAAGGACCGTAGCTGACGGGCAAGCTGGAACCGGTGTCTCCCCAGCCCGCCGAC 364
 |||||
 QY 156 ThrAspProGluLeu 160
 |||||
 Db 365 ACTGACCCCTGAGCTG 379
 |||||
 RESULT 51
 AAT93095
 ID AAT93095 standard; cDNA; 24379 BP.
 AC AAT93095;
 XX
 XX 11-MAY-1998 (first entry)
 DE Streptomyces frenolicin gene cluster.
 XX
 KW Frenolicin; antibiotic; feed additive; antioccidial; coccidiostatic;
 KW efflux pump; butyrate starter synthase; polyketide synthase; PKS;
 KW hemiketalase; ketoreductase; cyclase; dehydrase; ketoreductase;
 KW hydroxylase; Streptomyces roseofulvus; ds.
 XX
 OS Streptomyces sp.
 XX
 FH Key Location/Qualifiers
 FH CDS 636..2948
 FT /*tag= a
 FT /product= "80 kba non-membrane protein"
 FT /note= "gene A (specifically claimed)"
 FT 2945..3916
 FT /*tag= b
 FT /product= "membrane protein"
 FT /note= "gene B (specifically claimed)"
 FT 4020..4844
 FT /*tag= c
 FT /product= "protein with 6 membrane-spanning domains"
 FT /note= "gene C (specifically claimed)"
 FT 4841..6415
 FT /*tag= d
 FT /product= "ATP-binding component of ABC transporter"
 FT /note= "gene D (specifically claimed)"
 FT 6533..7183
 FT /*tag= e
 FT /product= "unknown non-membrane protein"
 FT /note= "gene E (specifically claimed)"
 FT 7344..8897
 FT /*tag= f
 FT /product= "putative efflux pump"
 FT /note= "gene F (specifically claimed)"
 FT 9164..10012
 FT /*tag= g
 FT /product= "transcription activator"
 FT /note= "gene G (specifically claimed)"
 FT complement(10105)..10621
 FT /*tag= h
 FT /product= "translationally coupled to gene I"
 FT /note= "gene H (specifically claimed)"
 FT complement(10618)..11628
 FT /*tag= i
 FT /product= "homologue of fabH"
 FT /note= "gene I (specifically claimed)"
 FT 11809..12066
 FT /*tag= j
 FT /product= "acyl carrier protein"
 FT /note= "gene J (specifically claimed)"
 FT complement(12154)..13209
 FT /*tag= k
 FT /product= "acyltransferase"
 FT /note= "gene K (specifically claimed)"
 FT 13409..14686
 FT /*tag= l

FT /product= "PKS ketoacylsynthase subunit"
 FT /note= "gene L (specifically claimed)"
 FT 14767..16047
 FT /*tag= m
 FT /product= "PKS chain length factor"
 FT /note= "gene M (specifically claimed)"
 FT 16120..16371
 FT /*tag= n
 FT /product= "acyl carrier protein used by the PKS"
 FT /note= "gene N (specifically claimed)"
 FT complement(16453)..16935
 FT /*tag= o
 FT /product= "putative hemiketal dehydrase"
 FT /note= "gene O (specifically claimed)"
 FT 17088..17903
 FT /*tag= p
 FT /product= "ketoreductase related to actIII"
 FT /note= "gene P (specifically claimed)"
 FT 17903..18898
 FT /*tag= q
 FT /product= "cyclase/dehydrase related to act VII"
 FT /note= "gene Q (specifically claimed)"
 FT 18895..19839
 FT /*tag= r
 FT /product= "cyclase/dehydrase related to actIV"
 FT /note= "gene R (specifically claimed)"
 FT complement(19990)..20907
 FT /*tag= s
 FT /product= "oxidoreductase"
 FT /note= "gene S (specifically claimed)"
 FT complement(20904)..22094
 FT /*tag= t
 FT /product= "quinone-forming hydroxylase"
 FT /note= "gene T (specifically claimed)"
 FT complement(22505)..22179
 FT /*tag= u
 FT /product= "not specified"
 FT /note= "gene U"
 FT
 PN EP806480-A2.
 PD 12-NOV-1997.
 XX
 XX 02-MAY-1997; 97EP-00107329.
 PF
 XX
 XX 07-MAY-1996; 96US-0016753P.
 PR
 XX (HOFF) HOFFMANN LA ROCHE & CO AG F.
 PA
 XX
 XX Reeves CD, Soliday CL;
 PI
 PI WPI; 1997-538619/50.
 DR P-PSDB; AAW34199, AAW34200, AAW34201, AAW34202, AAW34203, AAW34204,
 DR AAW34205, AAW34206, AAW34207, AAW34208, AAW34209, AAW34210, AAW34211,
 DR AAW34212, AAW34213, AAW34214, AAW34215, AAW34216, AAW34217, AAW34218,
 DR AAW34219.
 XX
 XX Streptomyces frenolicin gene cluster - useful for producing recombinant
 PT frenolicin antibiotics.
 PT
 XX Claim 1; Page 40-60; 66pp; English.
 PS
 XX This DNA sequence comprises the Streptomyces frenolicin gene cluster
 CC containing specifically claimed coding sequences (genes A-U) that
 CC respectively encode 21 proteins (see AAW34199-219) involved in frenolicin
 CC synthesis. The genes can be divided into 5 subclusters: (1) genes A, B,
 CC C, D and F encode an efflux pump; (2) genes H, I, J and K encode butyrate
 CC starter synthases; (3) genes L, M and N encode polyketide synthases (PKS)
 CC ; (4) genes O, P, Q and R encode a hemiketalase, a ketoreductase and
 CC cyclases/dehydrases; and (5) genes S and T encode a keto/enoyl reductase
 CC and a hydrolase. Also claimed are vectors, host cells (especially a
 CC Streptomyces sp., particularly Streptomyces roseofulvus), and the encoded
 CC proteins. Cells transformed using the above sequence can be cultured to

```
QY 2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlys 2259
Db 43508 CAGGTTACCGCCTT-----TGTCCACCGGAGCCGGCGGGA 43473
QY 2259 erLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerA 2279
Db 43472 CCCTGCCACCGCCGCGGTGGCGCGCGGTG----- 43438
QY 2279 snSerAlaMetValLysSerLysGlnGluLeuAsnLysLysLeuAsnThrHisAsnA 2299
Db 43438 ----- 43438
QY 2299 rgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaI 2319
Db 43437 ----- 43437
QY 2319 leThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrA 2339
Db 43387 ----- 43387
QY 2339 snMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluG 2359
Db 43386 ----- 43386
QY 2359 luSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProA 2379
Db 43367 CGGGCCACCTTTCGCGCGCGCGG-----CGGTTACCGCGGCGCGCGCGCGCGG 43314
QY 2379 laAlaMetProIleThrAlaAla-----AspGlyArgSerAspHisT 2393
Db 43313 CGGTGCCCGCGGTGAT-ATTGCCCGCGTTCGCGCGCGTCCGCGGACCGCGGACCA 43255
QY 2393 hrLeuThrSerProGlyGlyGlyLysAla-----LysValSerGlyA 2408
Db 43254 CCTTACACCGGACCGCGCGCGCGCGCGCGCTTGTTCGCGCGGTCCGCGGCG 43195
QY 2408 rgProSerSerArgLysAlaLysSerProAlaProGly-----LeuAla 2423
Db 43194 GCGCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 43135
QY 2423 erGlyAspArgProProSerValSerValHisSerGluGlyAspCysAsnArgArgT 2443
Db 43134 CGGTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 43076
QY 2443 hrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheP 2463
Db 43075 G-CGCGTG-----CGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 43041
QY 2463 roTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProG 2483
Db 43040 CAGCGGTGCCGTTA-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCG 43014
QY 2483 lyLeuProAlaGlySerGlyProLeuAlaGlyPro 2494
Db 43013 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 42979
RESULT 50
ACH20441
ID ACH20441 standard; cDNA; 381 BP.
XX
AC ACH20441;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult liver cDNA #53.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
```

```
XX
PD 17-APR-2003.
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 7653; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docID=20030073623
XX
SQ Sequence 381 BP; 72 A; 146 C; 105 G; 58 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.09e-12 Length: 381
Score: 645.00 Matches: 123
Percent Similarity: 98.40% Conservative: 0
Best Local Similarity: 98.40% Mismatches: 2
Query Match: 4.88% Indels: 0
DB: 8 Gaps: 0
US-09-522-753-5 (1-2517) x ACH20441 (1-381)
QY 36 AspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerPro 55
Db 5 GAGCTCGGGCTCTCTGGAGTACCAGCATCTCCCGGAGCTATGCTCCACCTGTCGCC 64
QY 56 GlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGlnProGly 75
Db 65 GGCTCCATCATCCAGCCCCAGCGGAGGCCCTCCCTGCTGTGTAGTTCCAGCCGGG 124
QY 76 AsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeu 95
Db 125 AATGAACGGTCCCGAGAGCTCCACCTGCGGCGAGAGTCCCACTCATACCTGCCGAGCTG 184
QY 96 GlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAsp 115
Db 185 GCGAAGTCAGAGATGGAGCTCATTTAAAGCAAGCGCCCTCGGCTAGAGTGTGCTGTGAC 244
QY 116 ProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeu 135
Db 245 CCCTGCTGCGACCGTCACCCCTGCTGTGACACGCGGCCCACTCGGGATCTGAAGACCTC 304
```

```
Db 44995 CCGGGCCACCATCGCGCGCTGGGCCCATCGGACTAGACCGCGCTGCGCCCTT 44936
Qy 1554 uProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLe 1574
Db 44935 GCCGCGCTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 44882
Qy 1574 uSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluLeuAla 1594
Db 44881 ACCACCGTCTCCGGCGCCACCGCGCTGCGCGGTCTCCCTTGC 44836
Qy 1594 sSerProHisSerThrValProGluHisHisProHisProHisSerProThrGluHis 1614
Db 44835 ----CCATTCCAGCCGCGTCCCGGTTCACACACACCG-- 44800
Qy 1614 uLeuArgGlyValSerGlyValAspLeuThrArgSerHisLeuProLeuAlaPheAspPr 1634
Db 44799 ---- 44783
Qy 1634 oThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaAlaAlaAlaAlaAla 1654
Db 44782 GCCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44752
Qy 1654 sLeuAlaProAsnProThrTyrProHisLeuTyrPro-----ProTyrLeuI 1670
Db 44751 -TTGCCGCGCGGTGCCACCTTGCCTGCGCGGTGTGCGCGCGCGCGCGCGCGCGCGCT 44693
Qy 1670 eArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleAsnAspTyrI 1690
Db 44692 CGGGGGCACTCCGGGTCGCGG-- 44671
Qy 1690 eThrSerGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLe 1710
Db 44671 ---- 44671
Qy 1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgG 1730
Db 44670 ----TCGCGG-- 44665
Qy 1730 yIleIleAspLeuSerGlnValProHisLeuProValLeu-----ValProProThrPr 1748
Db 44664 ----CCATTCCGCGGTGCGCGGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 44612
Qy 1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPh 1768
Db 44611 ATTGCTCCGCGCGCGCGGTGCGCGGTGCGCGCATTCGCCGCGCACCG----- 44560
Qy 1768 eSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPr 1788
Db 44559 ----CCGCTGCGCGCG-- 44540
Qy 1788 oThrThrThrSerSerSerGluArgGluArgAspArgGluArgAspArgAspAr 1808
Db 44539 GTCGCC-GCGCGATCCGCCAC-- 44519
Qy 1808 gGluArgGlySerIleLeuThrSerThrThrValGluHisAlaProIleTrpAr 1828
Db 44519 ---- 44519
Qy 1828 gProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlySerSe 1848
Db 44518 -CCGGGGTCCGCGAGCCCGCGTGCCTGCGCGGTGCGCGGTGCGCGCGCGCGCGCGCGCG 44466
Qy 1848 rSerArgProAlaSerHisSerHisAlaHisGlnHisSerProHisSerProHisThrGl 1868
Db 44465 CCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44406
Qy 1868 nAspAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db 44405 CGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44352
Qy 1888 rAlaVal-GluProSerLysProThr---ValLeuArgSerThrSerThrSerSerProV 1907
```


48671 Db -----CTGAGTCGCGTGTAGCGCGCTTTGCGCGCGTGCCG 48630
257 QY -----TyrAsnGlnProSerAspTh 263
48629 Db CCGGTGCGCGGTGCGCGCGTGAAGTTGGCCCGCGCGCTACCGCGCGCGCGCGGTG 48570
263 QY rArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMet---ArgLysIleLeuIle 282
48569 Db CCGCGGTGCGCGGTGCGCGCGGTGCGCGCTTGGCGGGTGGCGCGCGGTGCCT 48510
282 QY uTyrPheLysArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTy 302
48509 Db GCGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTCG 48450
302 QY rAspGlnLeuMetGluAlaLeuGluLysLysValGluAlaGluLeuAsnProArg 322
48449 Db CTGCGGTGCTGAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48390
322 QY rArgAlaLysGlnSerLysValArgGlnTyrTyrGluLysGlnPheProGluIleArgTy 342
48389 Db GCGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48336
342 QY sGlnArgGluGlnGlnGluArgMetGlnSerArgValGlnArgGlySerGlyLeuSe 362
48335 Db GCGCGCGCGGTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48290
362 QY rMetSerAlaAlaArgSerGluHisGluValSerGluIleAlaAspGlyLeuSerGlu 382
48289 Db -----GCTGGTGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48252
382 QY nGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAl 402
48251 Db TCACCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48227
402 QY aAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422
48227 Db -----48227
422 QY lTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPheArgGl 442
48226 Db -----CGACACCGCGGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 48177
442 QY uLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysTh 462
48176 Db CCGCGCGTTCG 48117
462 QY rValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLe 482
48116 Db GTGCGCGCGATGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 48057
482 QY uValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln 502
48056 Db TGTCGCGTGTGCGTGTAC 47997
502 QY nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 517
47996 Db CCG 47938
517 QY nGluGluLysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLys 537
47937 Db GCGCGCGCGTTCG 47878
537 QY uValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyLys 557
47877 Db GGTGCG 47824
557 QY pAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyAr 577
47823 Db -----GCGCGT 47777
577 QY gArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGlu-----593

47776 Db CCGCGCGACTGCTGTACCGCGCGCTTGC CGCGCGCGCGCTTCCACCGCGAACC CGG 47717
594 QY -GluAla-1IleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerS 613
47716 Db TCGCGCGGTGGCGCGCGGT 47657
613 QY rArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuGluGluHisGlyArgA 633
47656 Db CGCTTGGCGCGCGCG-CCGCGCGACACCGCGTTCCTGTGTGTGTGTGTGTGTGTGTGT 47598
633 QY snTrpSerAlaIleAlaArgMetValGlySerLysThrVal-----646
47597 Db GTGCTTCCGCGT 47539
647 QY -----SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspG 664
47538 Db GCGGTACCGCGGT 47489
664 QY lIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysL 684
47488 Db -----CGCACGACACCG-CCGCGCGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 47440
684 QY yLys-AlaProAlaAlaSerGluGluAlaAlaPheProProValValGluAspGlu 703
47439 Db GGAGCGT 47398
704 QY GluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAla 723
47397 Db -----GCCGCGCGCGCGCTCCACCGCGCGAACCC-----47368
724 QY LeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsn 743
47367 Db ---GGTGGCGCGGTGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 47320
744 QY AsnSerSerAspThrGluSerIle-ProSerProHisThrGluAlaAlaLysAspThrGl 763
47319 Db GGTCTCGCGTACCG 47260
763 QY yGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyPro-----ProProGl 781
47259 Db GATATCGCGCGGT 47200
781 QY yProProThrProProArgArgThrSerArgAlaProIleGluProThrProLaserGl 801
47199 Db TCGCGCGGTGCGCGCG-----TCACCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 47152
801 QY uAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProPro---Pr 820
47151 Db ATCG 47092
820 QY oValValProLysGluLysGluGluGluThrAlaAlaAlaProProValGluGlu 840
47091 Db GTTGT 47033
840 QY yGluGluGlnLysProProAlaGluGluLeuAlaValAspThrGlyLysAlaGlu 860
47032 Db CGAACCGGTGCGCGCGGT 46973
860 QY uProValLysSerGluCysThrGluGluAlaGluGluGluGluGluGluGluGluGlu 880
46972 Db CCGCGTTCACCGCGT 46916
880 QY aGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGlyLys 900
46915 Db T----GGCGT 4859
900 QY rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaTh 920
46858 Db CGACCGCGCGCTTACCG 46799
920 QY rCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer-P 940
46798 Db CGCGGT 46751

| | | | | |
|---|---------------|--|----------|------|
| QY | 2317 | -----ProAlaIleThrGly----- | -----Thr | 2322 |
| Db | 85205 | ACCACCCACCGCTTCGGCTACGGCCCGCTTCAGGGCGCTGGGGCCGCTGGCGCGCC | 85264 | |
| QY | 2323 | GlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly--- | 2341 | |
| Db | 85265 | GGCAGCGTGTGTACGCGGAGGTGCTCCCGAGTTCGCCACACGAGAGGGCGGCC | 85324 | |
| QY | 2342 | -----LeuGluAlaIleAileArgLysAlaLeuMet----- | 2351 | |
| Db | 85325 | TTGGGCTGCACCCGGCGCTCTGGACCGCGCTGCACGCGCGCTCTCGCGCAGCAG | 85384 | |
| QY | 2352 | -----GlyLysTyrAspGlnTrpGlu----- | 2359 | |
| Db | 85385 | CGCGACACCGAGCTCCCGTCTCTCGGAAGAGGTCACTCTGCACGCTCCGGCGCCACC | 85444 | |
| QY | 2360 | -----SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeu | 2377 | |
| Db | 85445 | GGCGTACGCTCGGCTCGCCCGACGGCCCAACGGCTGTCTGCACCGCGCGCAG | 85504 | |
| QY | 2378 | ProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerPro | 2397 | |
| Db | 85505 | CCGGCCGGCAACCCGTCGCCACCGTC-----ACCGCGCTGCTCGCGCGGCC | 85552 | |
| QY | 2398 | GlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerPro | 2417 | |
| Db | 85553 | CT-GGACGCGCAGCAGTGTACCATCCACAGCGCCCTGACCCGCGCAGCGCTCTTCACCT | 85611 | |
| QY | 2418 | Ala-----ProGlyLeuAlaSerGlyAspArgProProSerValSerValHisSer | 2435 | |
| Db | 85612 | GGACTGACCCCGCTCCGCTTCCGACACCGCACTCCGGCGCGCGCTCTCTCGG | 85671 | |
| QY | 2436 | GlulysAsp-----CysAsnArgArgThrProLeuThr | 2446 | |
| Db | 85672 | CCCGGACACCGCGCTGCTCGCGCAGCGCTCGGCGACCCGCGCGCTCGCACGCCAC | 85731 | |
| QY | 2447 | AsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPro | 2466 | |
| Db | 85732 | CCTCGA-----CGACCTCTCGCGCGGGACACACCCCGCGCGCGCTCTCT | 85779 | |
| QY | 2467 | Leu-----IleMetArgLeuGlnAlaGlyValMetAlaSerProProPro | 2481 | |
| Db | 85780 | CGTCCCTCGGCGCCCACTCGACGGCGACCGCGCAGCAGCGCAGC-CCCTCACCC | 85838 | |
| QY | 2482 | -----ProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAla | 2497 | |
| Db | 85839 | GCAGCGCGCTGACCTCTGTCAGCAGTGGCTCGCACCGACCGCTCGCGCAGCTCCCGCC | 85898 | |
| QY | 2498 | TrpAspGlu-----GluProLysPro | 2504 | |
| Db | 85899 | TGGTCTTCGTACCCACCGAGCGCTCGCGCA | 85928 | |
| RESULT 49 | | | | |
| AAI199682_39/c | | | | |
| Continuation (40 of 45) of AAI199682 from base 3900001 (Mycobacterium tuberculosis strain | | | | |
| WP Sequence split into 45 fragments LOCUS AAI199682 Accession Aai199682 | | | | |
| WP | Fragment Name | Begin | End | |
| WP | AAI199682_00 | 1 | 110000 | |
| WP | AAI199682_01 | 100001 | 210000 | |
| WP | AAI199682_02 | 200001 | 310000 | |
| WP | AAI199682_03 | 300001 | 410000 | |
| WP | AAI199682_04 | 400001 | 510000 | |
| WP | AAI199682_05 | 500001 | 610000 | |
| WP | AAI199682_06 | 600001 | 710000 | |
| WP | AAI199682_07 | 700001 | 810000 | |
| WP | AAI199682_08 | 800001 | 910000 | |
| WP | AAI199682_09 | 900001 | 1010000 | |
| WP | AAI199682_10 | 1000001 | 1110000 | |
| WP | AAI199682_11 | 1100001 | 1210000 | |
| WP | AAI199682_12 | 1200001 | 1310000 | |
| WP | AAI199682_13 | 1300001 | 1410000 | |
| WP | AAI199682_14 | 1400001 | 1510000 | |
| WP | AAI199682_15 | 1500001 | 1610000 | |

| | | | | |
|--|--------------|--|---------|-----|
| WP | AAI199682_16 | 1600001 | 1710000 | |
| WP | AAI199682_17 | 1700001 | 1810000 | |
| WP | AAI199682_18 | 1800001 | 1910000 | |
| WP | AAI199682_19 | 1900001 | 2010000 | |
| WP | AAI199682_20 | 2000001 | 2110000 | |
| WP | AAI199682_21 | 2100001 | 2210000 | |
| WP | AAI199682_22 | 2200001 | 2310000 | |
| WP | AAI199682_23 | 2300001 | 2410000 | |
| WP | AAI199682_24 | 2400001 | 2510000 | |
| WP | AAI199682_25 | 2500001 | 2610000 | |
| WP | AAI199682_26 | 2600001 | 2710000 | |
| WP | AAI199682_27 | 2700001 | 2810000 | |
| WP | AAI199682_28 | 2800001 | 2910000 | |
| WP | AAI199682_29 | 2900001 | 3010000 | |
| WP | AAI199682_30 | 3000001 | 3110000 | |
| WP | AAI199682_31 | 3100001 | 3210000 | |
| WP | AAI199682_32 | 3200001 | 3310000 | |
| WP | AAI199682_33 | 3300001 | 3410000 | |
| WP | AAI199682_34 | 3400001 | 3510000 | |
| WP | AAI199682_35 | 3500001 | 3610000 | |
| WP | AAI199682_36 | 3600001 | 3710000 | |
| WP | AAI199682_37 | 3700001 | 3810000 | |
| WP | AAI199682_38 | 3800001 | 3910000 | |
| WP | AAI199682_39 | 3900001 | 4010000 | |
| WP | AAI199682_40 | 4000001 | 4110000 | |
| WP | AAI199682_41 | 4100001 | 4210000 | |
| WP | AAI199682_42 | 4200001 | 4310000 | |
| WP | AAI199682_43 | 4300001 | 4410000 | |
| WP | AAI199682_44 | 4400001 | 4411529 | |
| Alignment Scores: | | | | |
| Pred. No.: | 1-51e-10 | Length: | 110000 | |
| Score: | 646.00 | Matches: | 563 | |
| Percent Similarity: | 29.92% | Conservative: | 226 | |
| Best Local Similarity: | 21.35% | Mismatches: | 966 | |
| Query Match: | 4.89% | Indels: | 894 | |
| DB: | 4 | Gaps: | 102 | |
| US-09-522-753-5 (1-2517) x AAI199682_39 (1-110000) | | | | |
| QY | 114 | ProAspProLeuLeuArgProSerProLeuLeu- | ----- | 124 |
| Db | 48983 | CCGAGCGCTGTGTGTCGGCGCGCACCGCTGCTACCGCGGCCCGCTTTCGCGCGGCC | 48924 | |
| QY | 125 | -----AlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeu | 141 | |
| Db | 48923 | CCGCCGCAAGCGGGCTGGCGCCCAATCCGGAACCGCTGCTCTTGGCGCGCTTA | 48864 | |
| QY | 142 | ThrGlyLysLeuGluProValSerProSerProProHisThrAspProGluLeuGlu | 161 | |
| Db | 48863 | -----CCGCTTGGCGCGCTCACCGCGGATG-----CCGCGTTGCCT | 48825 | |
| QY | 162 | LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArg | 181 | |
| Db | 48824 | GCGTGCCTGCGCGGTGTG----- | 48807 | |
| QY | 182 | GluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnLeuGlu | 201 | |
| Db | 48807 | ----- | 48807 | |
| QY | 202 | GluGluAlaLysProProGluProGluLysProValSerPro | 216 | |
| Db | 48806 | -----GCTAGGCGCGCGCGCGCTTGGCGCGCGCGCTGTGCGCGCGGTC | 48756 | |
| QY | 217 | ProProlleGluSerLysHisArg----- | 224 | |
| Db | 48755 | CCGCGCGGAGTGTGTCGCGCGCGCGCGCGCTTTCGCGCGGTGCGCGCGGTG | 48696 | |
| QY | 225 | -----SerLeuValGlnIleLeuTyrAspGluAsnArgLysLysAlaGluAla | 241 | |
| Db | 48695 | CCGCGCGGTGCGCTGTGTGGCGCTG----- | 48672 | |
| QY | 242 | HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro-Leu----- | 256 | |

Db 83086 GCTCGCTCGATCAAGTCCAAATCGGTCAACCCAGGAGCGCGGTGTGCGCGGTGT 83145
QY 1919 ----- 1919
Db 83146 CATCAAGATGGTCATGGCGATCGGCACGCGGTGTGCGCGACACCTGCTCGACGC 83205
QY 1920 -----ProL 1921
Db 83206 GCGGTCTCTCGACGTGATGGAGCGTGGCGCGCTCGAACTGCTCACCGAGCAGACCGC 83265
QY 1921 euGly-----GlyThrLeuaspGlyValTyrProThrLeuMetGluProVal 1937
Db 83266 CTGCGCGAGACCGCGCGCGCGTGTGCGCGGTGTCTCTCTCGGCATCAGCGGCAC 83325
QY 1937 euLeuPro-----LysGluAlaProArgValAlaArgProGluArgProArgAla 1954
Db 83326 CAAGCGCCAGCTGTCATCGAGCAGTCCCGCGCGGTCCCGCACGCGCGGTCCGC 83385
QY 1954 spThrGlyHisAla-----PheLeuAlaLysProProA 1965
Db 83386 CGACCGGTCTCGAGGACCGCGCGCGGTCTGCGCGCTCTGCGGCAAGACCCCGA 83445
QY 1965 laArgSerGlyLeuGlu-----ProAlaSerSerProSerLysGlySerGluProArg 1982
Db 83446 CGCCCTCCGCGACAGCGCGCGCGCTCTGCGCCACGTCGAGGCGCCACCCGCACTGCG 83505
QY 1983 -----ProLeuValProProValSerGlyHisAla 1993
Db 83506 CCCGTGCGACATCAGTACTCTCTGATCGCACCGCGACCGCT-----TCGA 83553
QY 1993 hrLeuAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProA 2013
Db 83554 CCACCGCGCGTCTCTCGGCACCGACCGCGCGAGGCGCTGCGCGCTCTACCGCGCT 83613
QY 2013 laProProAla----- 2016
Db 83614 CGCGCGCGCGAGACCGACCGCGCGCGCTCTACCGGACCGTCCGACCGCGCGCACCGC 83673
QY 2017 -----SerA 2018
Db 83674 CTTCCTCTTCGCGCGAGGCTCCCAACGGCTCGGCATCGGCGCGGTCTCTACGAGCG 83733
QY 2018 laSerAspProHis-----ArgGluLysThrGlnSerLysProPheSerileGlnGluLeu- 2036
Db 83734 GTTCCCGCGCTTCGCGAAGCCCTCGACACCGTCTCACCGCGCTCGACGCGAATCGG 83793
QY 2037 -----GluLeuArgSer-LeuGlyTyrHisGlySerSerTyrSerProGluGlyVal 2053
Db 83794 CCACCCCTCCGCGACATCATCTGGGCGGAGGACGCTCAACTCGTCGACCGGACCGGCTA 83853
QY 2054 GluProValSerProVal-----SerSerProSerLeuThr 2065
Db 83854 CACCCACCGCGCTGTTCGCCATCGAGTGGGACTCTTCGCGCTCTCGAAGCCTGGGG 83913
QY 2066 HisAspLysGlyLeu----- 2070
Db 83914 CATCACCGGACTTCGTGGCGCGCACTCCATCGGCGAGATCGCGCGCACACGTCGC 83973
QY 2070 ----- 2070
Db 83974 CGCGGTGCTCTCTCTCGGCGACCGCTGCGCGCTCTGTCGTGGCGCGCGCGTGTGATGCA 84033
QY 2070 ----- 2070
Db 84034 GTCGTCCCGAAGGCGCGGATGATCGCGGTCCAGGCCACCGAGGACGAGTCTTGCC 84093
QY 2071 ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro----- 2088
Db 84094 CCTCTCACCGACGACGTCTCGATCGCGCGCGTCAACAGACCGCGCGCTCGTGTGCTC 84153
QY 2089 -----LysGlnProGlyPro----- 2093
Db 84154 CGGCTACGAGACGCCACCTCGCGGTGCGCGCGCACTTCGCGGACCGAGGCGCGCGCAC 84213

QY 2094 -----ValLysLeuGlyGlyLeuAlaAla----- 2101
Db 84214 CACGCGGTGCGGTGTCAGCGACCGCTTCTCACCTCGCGCTGATGGCGCGATGCTCGACGA 84273
QY 2102 -----HisLeuProHis-----LeuArg 2107
Db 84274 CTTCCGCGCGCTGTCGAGAGCCTCACCTTACCGCGCGCGCGACACCCCGCTCTCTCAA 84333
QY 2108 Pro-----LeuProGluSer 2112
Db 84334 CTTGACCGGGAACCTGCGCGCGCGCGGTCTGCTGCGCGCACTACTGCTGGTCCGGCA 84393
QY 2113 GlnProSerSerSerProLeu----- 2119
Db 84394 CGTCCGCGGCGGTTCGCTTCCGCGACGCGCATCCGCTCGCGACCGCGCGGTCTAC 84453
QY 2120 -----LeuGlnThrAlaProGlyValLysGly 2128
Db 84454 CACTTTCGTGCAACTCGCGCGCGCGCGGTGTCTGCGCATGCGCGCAGGAGTCCGCGCC 84513
QY 2129 HisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr 2148
Db 84514 CGAAGCGCG----- 84522
QY 2149 ArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGly 2168
Db 84523 CGSCACCATCC-----GTCCTCGCGCGCGCGCGCGCGAGAACAGGCGCTCTCTGGC 84576
QY 2169 AlaSerCysPro-----ValLeuAspLeuArgArg 2178
Db 84577 CGCCCTCTGCGCACTCCGCGTCTCGCGGTGCGAGCGCGACTGCTGCGCCACCTTCCGCG 84636
QY 2179 Pro-----ProSerAspLeuTyrLeuPro----- 2186
Db 84637 CTTGACCGCGTCCGCGTGCACCTCGACCTCGACCTTCCAGCGCTTCCAGCAGCGTGTCTGGCC 84696
QY 2186 ----- 2186
Db 84697 CGCGCGCGACCGCGCGCGCGCGCGCGCTGCGCGCGCTGCGCGCGCGCGAACA 84756
QY 2186 ----- 2186
Db 84757 CCCTCTCTCGCGCGCGCGTGCACACTCCCGCGACGACGCGCGCACTCTTTCACCGCGCG 84816
QY 2187 -----ProProAspHisGlyAlaProAlaArg-----Gly 2196
Db 84817 CTTCTCTCGCGCACCGACCGCGGTGGCTGGCGGACACACCGTCTCTGGGCACTGCTGCT 84876
QY 2197 SerProHisSerGluGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216
Db 84877 CCGGCGCACCGCACTGTTGAACTCGCGTCCGCGCGGCGGACGAGA-CGCGCAGCGCGCC 84935
QY 2217 GlyGlyGluAspGlyLeuGluProValSerProProGluGlyMetThrGluProGlyHis 2236
Db 84936 ACCTGAAGAAGTCACTCACCTCGCGCGCGCGCTGACCTCCCGGAGGACGCGCGCACCTCC 84995
QY 2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256
Db 84996 TCCAGGTCCGCTCGGATCCG-----CCGACGACCGCGCGCGCC 85034
QY 2257 MetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThr 2276
Db 85035 GCACCGTCTACCGTCCACCG 85094
QY 2277 GluSerAsnSerAlaMetValLysSerLysGlnGluLeuAsnLysLysLeuAsnThr 2296
Db 85095 -----CCACCGGTGTGCTCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 85145
QY 2297 HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMet 2316
Db 85146 CGGTCTGCG 85204


```
QY 110 LeuGluLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPro 129
Db 77560 -----CGCCGACGGCGCTGCAAGCCTTCTCCGACCG 77592
QY 130 AlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys-----Leu 145
Db 77593 GGCAGGACGGACCGCTGGGGCGAGGGCGCGCATGCTGGTCTCATGGGCTCTCCGA 77652
QY 146 GluProValSerProProSerProHisThrAspProGluLeuLeuValProPro 165
Db 77653 CGCCAGCGCGAGGGCGCGCTCTCGCGTGTCTGGCGGCTCGCCATCAACAGGA 77712
QY 166 ArgLeuSerLysGluGluLeuLeuGlnAsnMetAspArgValAspArgGluLeuThrMet 185
Db 77713 CGG-----CGCCTCCAAAGCGGCTGACCGC----- 77736
QY 186 ValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluAlaAla 205
Db 77737 -----TCCACGGCGCTCCCG-AGCAGCGGTCA 77765
QY 206 LysProGluProGluLysProValSerProProIleGluSerLysHisArgSer 225
Db 77766 TCCGCGCGCGCTGACAGCGCCACCTCACCGCGCGGCATCGACGCGTCCGAGGCC 77825
QY 226 LeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeu 245
Db 77826 ACGGACCGGACACCTCGGCGACCCGATCGAGGCGCCAGCGCTCTCGCGCACCTACG 77885
QY 246 Glu-----GlyLeuGly-ProGlnValGluLeu---ProLeuTyrAsnGlnProSerAs 262
Db 77886 GACAGGACCGCGCGCGCTCTGGCTCGGCTCGGTGAAGTCAACATCGGCGCACACCC 77945
QY 262 pThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLe 282
Db 77946 AGGCGCG-----CTCGG----- 77958
QY 282 uTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyr 302
Db 77959 -----TGCCGCGCGCG 77969
QY 302 rAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgAr 322
Db 77970 TGATCAAAATCATCGCGTTCGAGCGCG-----CGTGC 78005
QY 322 gArgAlaLysGluSerLysValArgGluTyrTyrLysGlnPhe----- 337
Db 78006 TGCCGCGAGCTGCACGCCACCGAACCCACCGAGCTGCATGGACCGCGCGCTCCG 78065
QY 338 ----ProGluIleArgLysGlnArgGluLeuGlnGluArgMetGln-SerArgValGlyG 356
Db 78066 TCGACCTCTCGACGAGCGTTCGCTGGCGGAGACCGGACGCGCGCGCGCGCGCG 78125
QY 356 InArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleI 376
Db 78126 TCTCTCTCTCGGATCATGCGCGCACCAACGCG-----CAGGTCTATCTCTGAACAGG 78176
QY 376 leAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleP 396
Db 78177 CCCCCACCGCGCGAAGAGCCACCCAGAACCCCGCTCGCGCCCGC-----GCCGTGTC 78233
QY 396 roProMetLeu-----TyrAspAlaAspGlnGlnArgIleLysP 409
Db 78234 CGTGGCGCTCTCCGCGCGCAGCGCGCGCTCGACGCCCGCGCGCGCGCTCACCG 78293
QY 409 he-----IleAsnMetAsnGlyLeuMetAlaAspP 419
Db 78294 GCCACTCGCGACACCCCGGACCGCGACCCCTCGACGTGCGGTACGCGCTCGCGGACG 78353
QY 419 roMetLysValTyrLysAspArgGlnValMet----- 429
Db 78354 GACGCGCCACCTTCGAACACCGCGCGCTCTGCTCCCGACGGCACCGAACTCGGCCACG 78413
QY 430 -----AsnMetTrpSerGluGlnLysGlu----- 438
Db 78414 GAACCGCGCGGAGGCGCCCTGCGCGCTCTCTTCTCCGCGCAGGCTCTCCAGCGCGCGG 78473
QY 439 -----ThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleA 455
Db 78474 GCATGGGACGGAACTCCACGCCCGCTTCCCGGTGTTCCCGCGCGCTTCCGACGAGATCA 78533
QY 455 laSerPheLeuGluArg-----LysThrValAlaGluCysValLeuTyrTyrL 472
Db 78534 CAGCGCTCTCTGACACCCACCTCGACCGCGCGCTCGCGAG-----GTCTCTGGGG-CACC 78589
QY 472 euThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGlyL 492
Db 78590 GACGCGACCTCTGACGACACCGCTGGGCCCAACCGCGCTCTGTCGCGCTCGAGTTC 78649
QY 492 ysSer-----GlnGlnGlnGlnGlnG 499
Db 78650 GGCCTCTACCGCTGCTGCGTCCCTCGCGCTGACCCCGGACTTCTGCGCGCGCACCTCC 78709
QY 499 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArg----- 514
Db 78710 ATCGGCGAGCTCGCGCGCGCACGCTCGCGGGGTCTCTCTCTCGAAGACGCTGCACC 78769
QY 515 -----SerSerGlnGlnGluLysAspGluLysG 524
Db 78770 CTGCTCGCGCGCGCGCGCTCATGACGCGCTCTCGCGCGCGCGCGCGATGTCGCG 78829
QY 524 lu-LysGluLysGluAlaGluLysGluGluLysProGluValGluAlaAsnAspLysGlu 543
Db 78830 ATCCG-----GCCACGAGGACGAGGTACCCCGCCACCTCACCGACGAGCTCTCG 78880
QY 544 AspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAla 563
Db 78881 -----ATCGCGCGCTCAACGCGGCCACCTCGCTGCTGCTCGCGCGCGCACCGAGAAGCC 78934
QY 564 valAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThr 583
Db 78935 GTCCGCGCATCGGGCGCGC-----TTCACCGCGCGCAGGACCGCAAGACCCCGCGCTCGCG 78991
QY 584 ArgSerMetAlaAsnGluAlaAsnSerGluAlaIleThrProGlnGlnSerAlaGlu 603
Db 78992 GTACGACGCGCTTCCAC-----TCGCGCGCTCATGGACCGCATG 79030
QY 604 LeuAlaSerMetGluLeuAsnGluSerArgTrpThrGluGluMetGluThrAla 623
Db 79031 CTGGCG-----GAATTCGCGCGCGTC 79051
QY 624 LysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGly--- 642
Db 79052 GCCGCGCGCTGACCTACACGAGCGCGCATCCCGGTCTCTTCCAACTCACCGGACCC 79111
QY 643 ---SerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsn 661
Db 79112 GTCCGCGCGCTCGCGACCTGTCTCGCGCGCTACTGGGTCTCCCGACCGCTCGCGAGCG 79171
QY 662 LeuAspGluIleLeuGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArg 681
Db 79172 GTCCGCTTCCCGA-CGCGCTCACCGCCCTCACCGACCGCGCGGTGACACCGTCTGTCGA 79230
QY 682 LysLys-----LysLysAla-ProAlaAlaSerGluGlu-----Al 694
Db 79231 ACTCGCGCGCGAGCGGTGTCTCGCGCATGGCCGAGGATCTCTGCGGACGCGCGCGC 79290
QY 694 aAlaPheProProValValGluAspGluMetGluAlaSerGlyValSerGlyAsnG 714
Db 79291 CGCGCTCGCTGTGGCGCAAGGACCGCCCGAGGAGCTCTCGCGCGCTCACCGGCTCGG 79350
QY 714 uGluLysMetVal----- 718
Db 79351 CCGCGCGCACGTCCGCGCGCTCACGCTCGCGCGCGGCTCTTTCAGACGCGCGCGC 79410
QY 718 ----- 718
```


Db 21336 C- |||| : :
QY 1469 sLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProValHisPr 1489
Db 21348 GGGCCGGCCCTCGCC- - - - -GGTCGGCTCGACCTCGCCGTCCTGGCACCC 21398
QY 1489 o- - - - -LeuAspValMetAlaAspAlaAaGalLeuGluArgAlaCysTyrGluGl 1506
Db 21399 AGGCGACATCGCCCGCTGCTGGCGGCTGTATCGCACCCCGCTGGCGGCGACCGCG 21458
QY 1506 uSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAl 1526
Db 21459 CCAGG-TCTCGACAGCCCGGCGGCTCGCCAGCGGTTCGCGCCCTCGACGCGGCC 21517
QY 1526 aProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHi 1546
Db 21518 GCCGGC- - - - -GGGAAGCCCTCTCG- - - - -AACTC 21544
QY 1546 sGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProTh 1566
Db 21545 GTCGGCACCC- - - - -AGATCGCCAGGTCTCTCG- - - - -GCCACGGCGGACGCCA- 21587
QY 1566 rProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuTh 1586
Db 21588 -CCGAGGTGAGA- - - - -CGCGCGCCAGTTTCCAGGACCTCGGC 21625
QY 1586 rSerThrProArgGluIleAlaLysSerPro- - - - -HisSerThrVal-ProGluHisH 1604
Db 21626 TTGACTCCC- - - - -TCACGCGGTGAACCTCGCAACGCGCTGAACACC 21670
QY 1604 isProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuT 1624
Db 21671 GCCACCGGCTGGCGGTGCCGCCACCAT- - - - -GGTTCGACTAC 21712
QY 1624 yrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuA 1644
Db 21713 C-CGACACACACGCC- - - - -TCGCGGACCACTGC- - - - - 21743
QY 1644 spAlaAlaAlaAlaTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisL 1664
Db 21744 - - - - -CGCAGCACTCTCTGGGACCGAGCGGCGAGTCGACCA- - - - - 21779
QY 1664 euTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnT 1684
Db 21780 - - - - -CGCGGTCCCGTGGCCGACCCGTACCGCGGACCGACGACCCGATCGTCATCGT 21834
QY 1684 hrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetA 1704
Db 21835 CGGCATGGCTCGCGCTACCCCGGCGCATCGCTCACCCGAGGACCTCTGGCGCCTGGT 21894
QY 1704 laGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnT 1724
Db 21895 CAGCAGGGCGCGACCGCACTGGCCGCTGCCACCAACCGCGGT- - - - - 21941
QY 1724 yrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuV 1744
Db 21942 - - - - -GGGACCTGGACAACCTCTACGACCCCGACCCCGACCCCGCGCGG 21984
QY 1744 alProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrA 1764
Db 21985 CCGCACCCATCGCGCGCGCGCTTCTCTCGACGACCGCGGTCTCTTCGACCGCGACTT 22044
QY 1764 la- - - - -ProGlnProPheSerSerArgHisSerS 1774
Db 22045 CTTCCGGATAGCCCGCGGAGGGATGGCCACCGACTCCAGCAGGCGCTGTGTCTGA 22104
QY 1774 erSerProLeuSerProGlyGly- - - - -ProThrHisLeuThrLysProT 1789
Db 22105 ACTCTCTGGAGAGCCGTGAACCGCGCGCATCGACCCCGCTCACTGGCGGACTCGCG 22164
QY 1789 hrThrThrSerSer- - - - -Ser-GluArgGluArgAspArg 1802
: : : : : |||||

Db 22165 CACCGCGCTTTCGCGCGGTCTATGTACAAAGACTAGGACACACCTGACCGCGACGA 22224
QY 1803 GluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSerThrThrThrValGlu 1822
Db 22225 GTACGAGGCGTTCGCGGCAACGCGACGCGCCCGAGCGTCGCCTCGCGCGCGTCTCCTA 22284
QY 1823 HisAlaProIleThrArgProGlyThr- - - - - 1831
Db 22285 CACCTCTCGCTGGAAGGCGCGCGCTACGGTGGACACCGCTGCTCTTCTCTCTGGT 22344
QY 1832 GluGlnSerSerGlySerSerGlyGlyGly- - - - -Gly-Gl 1846
Db 22345 CGCCCTGCACGTGGCGCGCGCGGTTCGCGGGGGAGTGTCTGTGGCGTGGCGG 22404
QY 1846 ySer- - - - -SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIl 1863
Db 22405 TGGTGTGACGGTGTATCGACGCGGACGACGTTCTGTGGAGTTCTCGCGGACGGGGTCT 22464
QY 1863 eSerPro- - - - - 1865
Db 22465 GCGCCTGTATGTTCTGAAGGCGTTTCGCGAGCGCGCGGCGGTGGCTGCTCGA 22524
QY 1866 - - - - -ArgThrGlnAspAlaLeuGln- - - - - 1873
Db 22525 GGGCGTGGCATGCTGCTCTGAGCGCGAGTCCGACGCGGTGCGAACCGTCACGAT 22584
QY 1874 - - - - -ArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAl 1889
Db 22585 CTTGGCGGTGGCGGTTCGCGGTCAACAGGACGGTGCGTCCACCGTTCGACCGC 22644
QY 1889 aValGlu- - - - -ProSerLysProThrValLeuArg- - - - - 1899
Db 22645 GCCAACGCGCGCTCCAGCAGCGGTGATCCGTCAGGCGTTGGCCAGTGGCGGCTGTC 22704
QY 1900 - - - - -SerThrSerThrSerProVa 1907
Db 22705 CAGCGCGCAGTGGACCGCTTGGAGCGCACGCGGTACGCGTTCGGTACCCGAT 22764
QY 1907 l-ArgProAlaAlaThrPheProAla- - - - -ThrHisCys- 1919
Db 22765 CGAGGCGGAGCGCTCTCTGGCCACCTACGTCGCGACCGCGACCCCGAGAACCCGCTGCT 22824
QY 1919 - - - - - 1919
Db 22825 GTCGGCTCGATCAAGTCCAAACATCGGTCAACCCAGGACGCGCGGTGTCGCGGTGT 22884
QY 1919 - - - - - 1919
Db 22885 CATCAAGATGGTCAATGGCATGGCGCACGCGGTGCTGCCGACACCCCTGATGTCGACGC 22944
QY 1920 - - - - -ProL 1921
Db 22945 GCGTCTCTCGCAGCTCGATTGGAGCGTCGGCGCGTTCGAACCTGCTCACCGAGCAGACGC 23004
QY 1921 euGly- - - - -GlyThrLeuAspGlyValTyrProThrLeuMetGluProValL 1937
Db 23005 CTGGCGGAGACCGCGCGCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 23064
QY 1937 euLeuPro- - - - -LysGluAlaProArgValAlaArgProGluArgProArgAlaA 1954
Db 23065 CAACGCGCAGCTGTCATGAGCAGTCCCGGACCGCGCTCCCGCACCGCGCGCTCCG 23124
QY 1954 spThrGlyHisAla- - - - -PheLeuAlaLysProProA 1965
Db 23125 CGACCGGTTCGTCGAGAACCGCGCGCTCCCTGGGCGCTGTCGGGCAAGACCCCGA 23184
QY 1965 laArgSerGlyLeuGlu- - - - -ProAlaSerSerProSerLysGlySerGluProArg- - - - - 1982
Db 23185 CGCCCTCCGCGACGACCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 23244
QY 1983 - - - - -ProLeuValProProValSerGlyHisAlat 1993
: : : : : |||||

Db 23245 CCGCGTGGACATCAGCTACTCTCTGATGCGCACCGGACCGGCT- - - - -TCGA 23292

[illegible]

| | | | |
|----|-------|--|-------|
| Qy | 1116 | erValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValP | 1136 |
| Db | 20512 | CGGTTT-----TCGC | 20520 |
| Qy | 1136 | rCyTy-SerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProM | 1156 |
| Db | 20521 | CCTCCTTGACCTCGCCCGCACACGGCGCGGCCGACCCCGAGACCCTGGCCACCGCCCT | 20580 |
| Qy | 1156 | eTAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSerPro--- | 1174 |
| Db | 20581 | GCGCGCACCGACGACGACCGACTCGCGCTCGCGGCACCGACGTGACGCGCGCCG | 20640 |
| Qy | 1175 | -----ArgGlyGlnAlaGlyProProGluSerLeuGlyVal----- | 1186 |
| Db | 20641 | CTTGCGCCGTGTCGCTCGCCACCGAACCCACACCACTGGAAACCGGACGCGCATCTCT | 20700 |
| Qy | 1187 | ----ProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeuGlySerValProG | 1205 |
| Db | 20701 | GATCACCGCGGCACCGCGGCGCTGGCGGGGTCTCGCCGCCACCTGTGTGCGACCCCA | 20760 |
| Qy | 1205 | lyGlySerIleThrLysGlyIleProSerThrArgValPro---SerAspSerAla--- | 1222 |
| Db | 20761 | CGCGCTCGGCACCTGCTG-CTCGCAGCGCGCGCGCGCGCGCCGCGCACCGCGCCGACG | 20819 |
| Qy | 1223 | -----lleThrTyArgGlySerIleThrHisGlyThrProAlaaspVall | 1238 |
| Db | 20820 | ACCTGACGCGGAACCTCACCGGGCTCGGCGCCACCGTCACATCGCCGCTCGACGTCG | 20879 |
| Qy | 1238 | eutyrylsGlyThrIleThrArgIleIleGlyLuaspSerProSerArgLeuaspArgG | 1258 |
| Db | 20880 | CCGACCGCGCGCCCTCGCGCACTGCTCGGC----- | 20911 |
| Qy | 1258 | lyArgGluaspSerLeuProLysGlyHisValIleTyrgLuglyLysGlyHisValL | 1278 |
| Db | 20912 | -----ACGTCCTCGCGCGGCACCGCTC-----ACCGTCG | 20942 |
| Qy | 1278 | eusertyrGluGlyGlyMet-SerValThrGlnCysSerLysGluaspGlyArgSerSer | 1297 |
| Db | 20943 | TGCTCCACACCGCGCGCTGTCGACGACGGGTCTCTGGCTCCCTACCCCGCAGCGCC | 21002 |
| Qy | 1298 | SerGlyPro-ProHisGluThrAlaAlaProLysArgThrTyraSpMetMetGluGlyAr | 1317 |
| Db | 21003 | TGGACACCGTCT-----CGGCCCCAAGCGC | 21029 |
| Qy | 1317 | gvalGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPr | 1337 |
| Db | 21030 | A-----CGCGCTCGGCACCTGCA-----CGAGCGACGACCGCC | 21062 |
| Qy | 1337 | cLUArgHisSer-----ProHisileuLysGluGlnHisIleArgGlyse | 1354 |
| Db | 21063 | ACCTCGACCTGGACGCTTCTGCTCTTCTGTCGCTCGCGCCAC----- | 21108 |
| Qy | 1354 | rIleThrGlnGlyIleProArgSerTyrrValGluAlaGlnGluAspTyrrLeuArgGl | 1374 |
| Db | 21109 | -----CCTCG-----CAGCCCCGACAGGCAACTACGCGCGC | 21143 |
| Qy | 1374 | uAlaLysLeuLeuLysArgGluGlyThrProProProProProSerArgAspLeuth | 1394 |
| Db | 21144 | GCAAACGCTTCTCGACGC-----CCTCGCGCGCGCGCGCGCCACCG | 21188 |
| Qy | 1394 | rGluAlatyrylsththrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLe | 1414 |
| Db | 21189 | GCCTGCGCGCACCTCTCGCTCGCTGGGCGCGTGGACCCGACGCGCGCGCCACCG | 21243 |
| Qy | 1414 | uValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGlu---- | 1432 |
| Db | 21244 | -----AAGCAGCCTGTCCA-----CCTCGACGTGACGCGCA | 21275 |
| Qy | 1433 | -----LeuArgHisThrProGluLeuPro-----LeuAlaProArgProLeuLysGluGl | 1449 |
| Db | 21276 | TGCGCCGCTCGGCATGCCCGCGCTGACCTCTGGAAAGGCGACCGCCCTCTTCGACGCGG | 21339 |
| Qy | 1449 | ySerIleThrGlnGlyThrProLeumystyrAspThrGlyAlaSerThrThrClySerLy | 1469 |


```
Db 10251 CCACACAGACCGCGGACCCCGCTCCGGCTCCGGCTCCGGACGCGCTCTCTCCA 10192
Qy 2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuArgSerLeuGly 2042
Db 10191 CGATCATCGCATTCGTCCACTGATCCGGAACGACGACACACAGCCCGCGCACCC 10132
Qy 2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerPro 2062
Db 10131 GCCACCCCGCGACCGCGCGCTCCGCTATGACCGATGTTCCGACTTCACCG 10078
Qy 2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeu 2082
Db 10077 ACCAGTCCACAT-----GGCGACGGCTCATCCA 10048
Qy 2083 GluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHis 2102
Db 10047 CATGCAACGCTTCGCGACACACACCGCGCGCCCGCGCA---GCACCATCTTCATCACAC 9991
Qy 2103 LeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThr 2122
Db 9990 CAGCCACACCGCGCGCGCTGCGTATGACCGATGTTCCGACTTCACCG---ACCCCAACA 9934
Qy 2123 AlaProGlyValLysGlyHisGlnArgValThrLeuAlaGlnHisLeuSerGluVal 2142
Db 9933 CCACCG-----9928
Qy 2143 IleThrGlnAspTyrThrArgHisHisProGlnLeuSerAlaProLeuProAlaPro 2162
Db 9927 -----GACGCCACCGCAGCGCCCTGCCATAGTCCGCAACACGCTGAGCCT 9877
Qy 2163 LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp 2182
Db 9876 CGATCGGATCACCAACCGCGTCC-----CCGTACCATGCGCTCCACCAT 9829
Qy 2183 LeuTyrLeuProProProAsp-----HisGlyAlaProAlaArgGly 2196
Db 9828 CCACATCAGCAGAACACAGACCCCGCTCAGCCACCGCGCAGCATCACCCTGTGTGG 9769
Qy 2197 SerProHisSerGluGlyGlyLysArgSer-----2206
Db 9768 ACGGACCATTCGCGCGCGTCAACCCATTCCAGCACCATCTGATTGACCGCATCCCC 9709
Qy 2207 ----ProGluProAnLysThrSerValLeuGlyGlyGluAspGlyIleGluProVal 2225
Db 9708 GCACACCGCAACACCCGATGACCAACCGCGCGCATCCGACACCGCTCCACCAACA 9649
Qy 2226 SerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu 2245
Db 9648 ACACACCCACACCC-TCGGCCCCAGCGGTGCGCTCCGCGCGCGG---AAGCGCTTG 9593
Qy 2246 TyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer---ProGlyAsn 2264
Db 9592 CAGCGCGCTCTCGGAGAGCGCCCTTGCCTGCGTGAAGTCCAGCAATGCGCGCGAA 9533
Qy 2265 -----ThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMet 2282
Db 9532 GGCATCACCGCTGCTCGCGCGCAATGCCAGTGAGCACTCGCCACCGCGCAGCGCTGC 9473
Qy 2283 Val-----LysSerLysLysGlnGluLeuAsnLysLysLeuAsnThrHisAsn 2298
Db 9472 ACCGCCAGGTGCATGCCACGACGACGACACACGCCGTGTCACCGTGACCGGAGGC 9413
Qy 2299 ArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAla 2318
Db 9412 CCCTCCAGACGAGGAGTAGCGCATGCGCGCGGAGCCACGCTGGCGGTATTGCCGGTC 9353
Qy 2319 IleThr-----2320
Db 9352 AGGACGTAGCGCGCGCGAGTGTGTCGCGATCTGTCAGTGGGGCGCGTAGTCTGCGCT 9293
Qy 2321 -----GlyThrGlyLeuMetThrTyr 2327
Db 9292 GTGGCTCCGATGTACACGCGCGGTGTGCTGCTCCGCGCAGACTGCGCGGGCTCACGCGCGCG 9233
```

```
Qy 2328 ArgSerGlnAlaValGlnGlnHisAlaSer-----ThrAsn 2339
Db 9232 CGTTGAGGGCTCCAGGACGCTCCAGCAGCAACCGCTGCTGCGGGTCCATCACCAAC 9173
Qy 2340 MetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu 2359
Db 9172 GCCTACGCGCGGAGATACCGAAGAAGGCGCATCGAACTCGCCGCGCTGTACAGAAAG 9113
Qy 2360 SerProProLeuSerAlaAsnAla-----PheAsnProLeuAsnAlaSerLeu 2377
Db 9112 CTCCCGCGCGGTGCGGAGTCAGTTCACGCTCCGCTCCGTTGGGGAAGCCGGAC 9053
Qy 2378 ProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerPro 2397
Db 9052 ACGCGTCTCGCTCGCGGAGCGGCGCACAGTCTCCGAGAGGTCACGCCACCC 8993
Qy 2398 Gly-----GlyGlyGlyLysAla 2403
Db 8992 GGAACGCGCACGCCATGCCAGATGGCGATCGGCTCGTCCGGGTCGGGCTGGC 8933
Qy 2404 LysValSerGlyArgPro-----SerSerArgLysAlaLysSerPro--- 2417
Db 8932 TCCGTGCGCGGTGCGCTGCTCTTCCCTGCTTCCGGGGCGCTCGTCTGCTGCCCTGG 8873
Qy 2418 -----AlaProGlyLeuAla-----SerGlyAsp 2425
Db 8872 AGGAACGCTGCCAGTCCGCGCGGCTCGGTGTCGAAGAGCACGCTGCTGGCAGCGG 8813
Qy 2426 ArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeu 2445
Db 8812 AGCCCGAGTCCGCTGTCGAGC-----CGTCCCGCAGT 8780
Qy 2446 ThrAsnArgValTrpGluAspArgProSerSer-----AlaGlySer-ThrPr 2461
Db 8779 TCCAGCGCGCTGACGAGTCGAAGCCGAGTCCGGAACCGCGGTCGGGTGATCCGC 8720
Qy 2461 oPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProPr 2481
Db 8719 TCGGCCCGGGAAT-----GTCCGAGCACAGCCG--- 8692
Qy 2481 oProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGlu 2501
Db 8691 -----CGCGTGGGAGCGCACGAGGTGCAGACGTCGTGCGCGCGCGCGGAGC 8642
Qy 2501 uProLysProLeuLeuCys 2507
Db 8641 GCCTCGTCCGTCCCGTGC 8623

RESULT 47
AADI17184
ID AADI17184 standard; DNA; 65140 BP.
XX AC AADI17184;
XX DT 29-NOV-2001 (first entry)
XX ST Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
DE Streptomyces noursei.
XX KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; nysI; de.
XX OS Streptomyces noursei.
FH Key Location/Qualifiers
FT CDS complement(1..1035)
FT FT /*tag= a
FT FT /product= "NysD2 partial protein"
FT FT /note= "CDS does not include stop codon"
FT FT complement(1056..2576)
FT FT /*tag= b
FT FT /product= "NysD1 protein"
FT FT 2806..6906
```

Qy 1543 yrGluAsp-----HisGlyAla-----ProPheAlaGlyHisL 1554
Db 12366 --CCGACGGCTGATGATACCAAGTCTCGGCCCTTCCCGCCAGCAGCGGGCGAC 12309
Qy 1554 euProArgGlySerProValThrMetArgGluProThrPro----- 1567
Db 12308 GTGGCGCCAGCGCCCGGCTCTCGGTCCACGAGCAGCGTGGCGGACCACTC 12249
Qy 1568 -----ArgL 1569
Db 12248 GCGCCCGCCGCTGGCTCGACATACGCTGACCGCGCTACGAACGAGCGCTCGGGCGAG 12189
Qy 1569 euGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArg---LysLeuThrSerT 1588
Db 12188 G-ACGGAGAGGCGCGTACGCAAGCTGGTCTCGTCCGCGAGCGCCGCCACGACATCGA 12130
Qy 1588 hrProArgGluAlaLysSerPro-----HisSerThrValProGluHis- 1603
Db 12129 CCAGCGTCCGCGCGCTCGTCCAGTGTTCGGGAGGTGCACCAACCGCCGCATA 12070
Qy 1603 ----- 1603
Db 12069 TCTCGCGCTGTTCAGGCGCAGACCCGCTCCAGTCCCGAGCAGCGCTGCTGGGTC 12010
Qy 1604 -----HisProHisProIleSerProTyrGluH 1613
Db 12009 GCTCGGTGACTACCGGGCAGTCCGCTACGCGGCTCGGTGCG- GAACCATACGGGC 11951
Qy 1613 isLeuLeuArgGlyVal----- 1618
Db 11950 ACCTCGCGCGCGTTCGGCATCGCTGACGAGCGCGAGGAGCGGCATCGCGACG 11891
Qy 1619 -----SerG 1620
Db 11890 GSCAGCTCGGGTGTCTACCGTGGCGGCTTCTGTCGAGGGCCAAAGAGGAGCACACCG 11831
Qy 1620 lyValAspLeuTyrArgSerHisIleProLeuAlaPheAsp-----ProThrSerI 1637
Db 11830 GCGAGTCTGCTGCGCCAGCCGCTTCGCGAGAGCTCGACAGCGCTTCGCGTCTGG 11771
Qy 1637 lePro----- 1638
Db 11770 TCCGCTCATCGACCAACGCGGACGCGCGCGCTCGCTCGAGTCTCGATG 11711
Qy 1638 ----- 1638
Db 11710 GCGCGCAACAGTAGTCTACGCGCCAGAGCTGCGCGAATGCGCAGCGCCAGACCG 11651
Qy 1639 -----ArgGlyIleProLeuAspAlaAlaAlaTyrT 1650
Db 11650 GACAGCGTGGGAAGTGTGTCGTGACGGGGCGCCACTCGACGTGTAGCGCAGTCC 11591
Qy 1650 yrLeuPro-----ArgHisLeuAlaProAsnProThrTyrProHisLeuT 1665
Db 11590 TCCAGCGCTCGCTTCCTGTCGCGCCCTTGTGCGAGCCAGTA-----GCGT 11540
Qy 1665 yrProProTyrLeuIleArgGlyTyrProAspThrAla-----AlaLeuGluAsnA 1682
Db 11539 TGCGTTTGGAGGGGTAGGTGGGAGGTGCGACGTGCGGGGCTCGCACCTCGAAGCG 11480
Qy 1682 rgGlnThrIleIleAsnAspTyrIleThrSerGln----- 1693
Db 11479 CGGGNACAGTCGACCGGACGCCCTTGA- CGAAGCTTCCGCGAGGAGGTGAGAACCG 11421
Qy 1694 -----GlnMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeu- 1710
Db 11420 ATCCAGCGCGCTTCGTTCAGCGCGAGCGAGTGCAGAGCTCGCGTCTCGAAGGT 11361
Qy 1711 -----ArgGlyLeu----- 1713
Db 11360 GTCGAGTGTTCGTGATGCGCGGAGCAGAGCGGGGTGCGGCTGCACTCGACGAAGCG 11301
Qy 1714 ----SerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleI 1732

Db 11300 GCCATCCCTGAGCCAGCAGCGCTTCTTGAACCGGACCGCTCTCCGCGAG 11241
Qy 1732 leAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProA 1752
Db 11240 ATTCTGTAACCACTACCCCGCATCCAGAGCGGTGCTGCAGCAGTCCGCGCTCACCGT 11181
Qy 1752 laThrAlaMetAspArgLeuAlaTyr-----LeuPro----- 1762
Db 11180 CGAAGAAGCGTACGTTCCCGATGTAGGACGATCGTCCGCGCTTCCAGCAGATC 11121
Qy 1763 -----ThrAlaProGlnProPheSerSerArgHisSer-----SerSerProLeuSerP 1779
Db 11120 ATCGCGTAACCTGCTCCATCTCGCGGAGTGGAGGATAGTCCACCGGAGCAGCAGC 11061
Qy 1779 roGlyGlyProThrHisLeuThrLysProThrThrSerSerSerGluArgGlu-Arg 1798
Db 11060 CCGGACACCTCCCGCTCAACACCGGCCAACACTATCCAGGGCTGGAGCTGACCCGA 11001
Qy 1799 AspArgAspArgGluArgAspArgGluArgGluArgGluSerIleLeuThrSerThr 1818
Db 11000 CACCAGTTCGAGCAGCGGCTTACCGCGC-----AACGGAAC 10959
Qy 1819 ThrThrValGluHisAlaProIleTrp-----ArgPro-----GlyThrGluGln 1833
Db 10958 CTTGCGCTCGAATCT-CCTCCAGCATGTGTGCGGACACGCGCGCGCGAGCTGACGGAGA 10900
Qy 1834 SerSerGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1853
Db 10899 CCATGCGCGCCACACGAGCATCCGGCGATCGCGCGCTCCGCAACACACACACCGCG 10840
Qy 1854 His-----SerHisAlaHisGlnHisSerProIleSerProArgThrGlnAsp 1869
Db 10839 CACCATCGGCAGACTCAACCCCGCAGCACCAAGCGCAGCAATCTCACTCCCTGCGAAT 10780
Qy 1870 AlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla 1889
Db 10779 GCGCCACACCGCAGCAGGCTCCACACCGTAATACCGCAGGTCCGCGCGCAGTGACACCA 10720
Qy 1890 ValGluProSerLysProThrValLeuArgSerThrSerThrSerProValArgPro 1909
Db 10719 TCACCGCCCAACGCGCGTGCACCATCAACCCCGCCCAAGACAGCTTCACCCAC 10660
Qy 1910 AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThr----- 1924
Db 10659 CGCGCAACACTCGACCGCAGCAGCAACCGGTCAACGGTCCAGCAGCGCAGCTCCG 10600
Qy 1925 -----LeuAspGlyValTyrProThrLeuMet----- 1933
Db 10599 CCACCGCTCCGGAACACCGCGCAGCATCCAGCAACCCAGCGCATCCCCACCCT 10540
Qy 1934 -----GluProValLeu 1937
Db 10539 GCGAACCTGACAGGAAGAACGACACACACGACCCCTCAGCGGAGCCACACCG 10480
Qy 1938 LeuProLysGluAla-----ProArgValAlaArgProGluArgProArgAlaAsp 1954
Db 10479 ACACCCCGCGCACAAACCCCGCAGCGAAGCGCGCAGCACTCCGCCATGGCAT 10420
Qy 1955 ThrGlyHisAlaPheLeuAlaLysProProAlaArg-----SerGlyLeuGluProAla 1972
Db 10419 CGAGTTCACCTGCGACCATCATCCGCGGTGCTCGAACCCGACCTCGTGCCACCAACG 10360
Qy 1973 SerSerProSerLys-----GlySerGluProArgProLeuValProProVal 1988
Db 10359 ACCAGCCCATCATGACCGCTTACCCGACAGCTGCGGAGCCAGCGTGCCT--- 10303
Qy 1989 SerGlyHisAlaThr-----IleAlaArgThrProAlaLysAsnLeuAlaProHis 2005
Db 10302 --GCTCAGCAACCGCCCGCATCCCGCGGACACACCC-----ACGCGCACACAC 10252
Qy 2006 ---HisAlaSerProAspPro---ProAlaPro---ProAlaSerAlaSerAspProHis 2022

Db 14412 GACGCCACTGTGACGACCTGCCCATACCTGCCAGCAACGCTGAGCCTCGATCGGAT 14353
QY 989 AlaAlaProThrLysProAlaProAlaProProAlaProProGlnAenLeuGlnProGlu 1008
Db 14352 CACCCAACCGGTCCCGTGCATGCGCTCCACCATCCATCCGACGAGGACGAC 14293
QY 1009 SerAspAlaProGlnGlnProGlnGlnSerSerProArgGlyLysSerArgSerProAlaPro 1028
Db 14292 -----CCGGCTCAGCCA-----ACGCCGACGGA-----TCACCGCTGCT 14257
QY 1029 ProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCys 1048
Db 14256 GCACGCGACCATTCGGCG-----CGCGAGACCAT 14227
QY 1049 TrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerProHis 1068
Db 14226 TGGACGCACCATCTGATTACCGCACTCCCC-----GCACCAACGCCA 14182
QY 1069 AlaPro---AspProSerAlaPhe-----SerTyAlaProPro---GlyHisProLeu 1084
Db 14181 ACACCCGATACCCCAACCGCGCGCATCCGACCAACCGCTCCACCAACACACCCACAC 14122
QY 1085 ProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsn 1104
Db 14121 CCT-----CGGACGACCAAAACCATCCCGCAAGCCGCAACGCTTGCACCGCCAT 14068
QY 1105 Pro-----ProLeuIleSerSer 1111
Db 14067 CCACGCAAAACCCGCTCGGACTGACGTGCTCAAGCTACCGGGGACGACATACCG 14008
QY 1112 AlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVal 1131
Db 14007 TCACCCCAACCCCAACG-----CAAGCG 13984
QY 1132 GlnLeuHisValProTySerGluHisAlaLysAlaProValGlyProValThrMetGly 1151
Db 13983 AACACTCACCCCGCGCAACCTTGACCCCGAGATGCAACGCCAACCAACGACGACAAAC 13924
QY 1152 LeuProLeuProMetAspPro-LysLysLeuAlaProPheSerGlyValLysGlnGlu 1171
Db 13923 ACGGCTATCCCGTACCGTCACCGAGGCCCTC-----AAGACCGAAGC 13882
QY 1171 nLeuSer-----ProArgGlyGlnAla-----GlyProProGluSerLeuGlyValPr 1187
Db 13881 TGTAAGCAACACGCCCCGAGCAGCGCTGGAGATGGAGCGGTGAGCAGGTAGCCCTCCA 13822
QY 1187 oThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu-GlySerValProGlyGlyS 1207
Db 13821 GCTCGGACGGACGTCGCGCCGCTGGCGCGGTAGTCTGTTGACATCCCGAAGAAC 13762
QY 1207 erIleThrLysGlyIleProSer-----ThrArgValP 1218
Db 13761 CA-----CGGTCGAGTCCCGCCAGCTTGAAGAGATGATGCGCGCCCGTTCGAGGCGCT 13705
QY 1218 roSerAspSerAlaIleThrTyArg-----GlySerIleThrHisGlyThrProAla 1236
Db 13704 CCCAGGACGCTCTCAGCAGCAACCGCTGCTGCGGGTCCATCGCAACGCTCA----- 13652
QY 1236 spValLeuTyLysGlyThrIleThrArgIleIleGlyGluAspSerProSer-----A 1254
Db 13651 -----CGCGCGAGATACCGAAGAGCGCGCTCGAAGCTCCCGCGCTCGTGCA 13603
QY 1254 rgLeuAspArgGlyAtgGluAspSerLeuPro-----LysGlyHisValIleTyGluG 1272
Db 13602 GGAAGCACCTCAGCATATGATGACGTACCGCGCGGTCCGATCGGATCGTACAGGC 13543
QY 1272 lyLysLysGlyHisValLeuSerTyGluGlyGlyMetSerValThrGlnCysSerLysG 1292
Db 13542 GCCCAAGTCCCAACCCCGGTTCGCGGAGTCCGAGACAAACGCTCTCACGAGCGGTGA 13483
QY 1292 luAspGlyArgSerSerSerGly-----ProProHisGluThrAlaAlaProLysA 1309
Db 13482 GCACGTCCACAGATCTCTCCGGGGAAGCGATGCTCCCGG-GAAGCGGCATGCCATTCCC 13424

QY 1309 rgThrTyAspMetMetGluGlyArgVal-----GlyArgAlaIleSerSerA 1325
Db 13423 ACGATCGCATCGGCTCGTCGATCGTCTTCCACGCGCGCAAGCACCCTGGTCA 13364
QY 1325 laSerIleGluGlyLeuMetGlyArgAlaIleProGluArgHisSerProHisIle 1345
Db 13363 CCATCTCTCGATCAGCTCGGAGCGCAGATGGCCGCGCCGCGGCTTCGGGTAG 13304
QY 1345 euLysGluGlnHisIleArgGlySerIleThrGlnGlyIleProArgSerTyValG 1365
Db 13303 TCGAAGACGCGCTCAGCGGCAGTCTGAGCCGCTGGCGGTGT-----CAGCCGG 13253
QY 1365 luAlaGlnGluAspTyLeuArgArg---GluAlaLysLeuLeuLysArgGluGlyThrP 1384
Db 13252 TTGCGCAGTTTCAGCCGAGGTTCAGCGAGTCCGATCCGAGATCTTGAACGCTC-GGTCCGG 13194
QY 1384 roProPro-----ProProProSerArgAspL 1393
Db 13193 AGCCACCGCTCGGCGCCGGAATGTCCGAGGACGACGCGGCTGCTCCCTCAGAGCTG 13134
QY 1393 euThrGlu----- 1395
Db 13133 CACCAAGAGAGTTCCTGCTCGACCCGCTGAGCCCGGAGACGACGCAAGCTCGGG 13074
QY 1395 ----- 1395
Db 13073 TTCGCACGCTCGGCTCGAGGTTCGCGGGCTGAGCCGACTCCAGCATCGCGGTGC 13014
QY 1396 -----AlaTyLysThrGlnAlaLeu-----G 1403
Db 13013 CTGGGTATCTCTCCGATCAAGGGCGGTGCGCCAGTGTGTACGCCAGGAAGAACCG 12954
QY 1403 lyProLeuLysLeuLysPro-AlaHisGluGlyLeuValAlaThrValLysGluAlaGly 1422
Db 12953 GCTCAGTCAATGTCGCCACGCCCACTGCGGTTTCTGCTCGCAGCGGTGGTGCAG 12894
QY 1423 ArgSer-----IleHisGluIleProArgGluGluLeuArg 1434
Db 12893 CGCAGCGATGGCGCGCGGTTCCATGGGAGTACACGTCGCTGGAGCGCTTCGCC 12834
QY 1435 HisThrProGluLeu----- 1439
Db 12833 CACCGACCTTTCATGTGCCATCGCGCGGACCGGTCGCCAGCGATGGAGGT 12774
QY 1440 -----ProLeuAlaProArgProLeuLysGluGlySerIleThrGln----- 1453
Db 12773 CGCACGAGCCCTGAGCCCTGCGCCATTCGCGAAGGGGTCCCAAGAACAGTTTCCCGG 12714
QY 1454 -----GlyThrProLeuLysTyAspThrGlyAlaSerThrThrGlySerLysLys 1470
Db 12713 CGGTTAGTTGGCTGCC-----GGGATTCGAGGTGGCGAGATGACGAGAACAG 12660
QY 1471 HisAspValArgSerLeuIleGlySerProGly----- 1481
Db 12659 TACGAGGCGGAGAG---GTCGAGGTCCCGGGTCAGTTGTCGAGGTTCAGGTGGCGGT 12603
QY 1482 -----ArgThrPheProProValHisProLeuAspVal----- 1492
Db 12602 GACTTTCGCTCGCAGGCGGTGTTCATCTGCTCCGTTGTGAGCACCGCTGCTATGCCGTC 12543
QY 1493 -----MetalAspAlaArgAlaLeuGluArg-----AlaCys 1503
Db 12542 GTCCAGAGCACCGCGCGGTGGATGACGGCGGTGAGTCCGCGCGGATCCGTGCGCAGCA 12483
QY 1504 TyrGluGluSerLeuLysSerArgPro-GlyThrAlaSerSerSerGly-GlySerIleA 1523
Db 12482 GTCGAGGAGCGTCTCCAGTGCCTGCGGCTCGCTGATTCACGCGCGGCGGCTCACCCG 12423
QY 1523 laArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrT 1543
Db 12422 GACACCCATGCTCGGTAGGCTGTCCGAGGGGAGCGGCGGCTCGCTCGCTC----- 12367

XX WPI; 2000-365602/31.
 DR P-PSDB; AAY92707, AAY92708, AAY92709.
 XX
 PT Recombinant DNA compound encoding oleandrolide polyketide synthase for
 PT synthesizing polyketides comprising a coding sequence for a domain of a
 PT loading module or any one of extender modules.
 XX
 XX Disclosure; Page 14-26; 86pp; English.
 PS
 XX This is part of the Streptococcus antibioticus oleandomycin gene cluster.
 CC The oleandrolide polyketide synthase (PKS), also known as 8,8a-
 CC deoxyoleandrolide synthase, is encoded by three open reading frames (ORF),
 CC designated oleA, oleAII and oleAIII. The PKS is a type I "modular"
 CC enzyme, where each ORF encodes 2 extender modules and the first ORF also
 CC encodes the loading module. Each module is composed of at least a
 CC ketosynthase (KS), acyl-transferase (AT) and an acyl carrier protein
 CC (ACP) domain. The oleandrolide PKS loading module contains an inactivated
 CC KS, called KS-Q, where Q is the abbreviation for glutamine, present
 CC instead of the active site cysteine required for activity. The large
 CC multifunctional PKS enzymes catalyze the biosynthesis of polyketide
 CC macrolactones through multistep pathways involving decarboxylative
 CC condensations between acylthioesters followed by cycles of varying beta-
 CC carbon processing activities. The macrolide product of the PKS, 8,8a-
 CC deoxyoleandrolide, is further modified by epoxidation and glycosylation to
 CC yield oleandomycin, an antibacterial polyketide. The invention concerns
 CC an isolated recombinant DNA compound, comprising a coding sequence for a
 CC domain of loading module or any one of extender modules 1-4 or 1-6,
 CC including an oleandrolide PKS operably linked to a promoter. Also
 CC discussed are recombinant oleandrolide PKS in which the module 1 KS domain
 CC is inactivated by deletion or other mutation. In particular, the
 CC inactivation is mediated by a change in the KS domain that renders it
 CC incapable of binding substrate (the KS1-o mutation), rendered by mutation
 CC in the codon for the active site cysteine. The oleandrolide PKS is useful
 CC for synthesizing polyketides, which are useful as antibiotics and
 CC motilides. Heterologous expression of oleandrolide PKS in host cells such
 CC as Streptomyces coelicolor and S. lividans is also made possible.
 CC Unmodified oleandrolide compounds can be provided to cultures of
 CC Saccharopolyspora erythraea and converted to the corresponding
 CC derivatives of erythromycins A-D. (Updated on 06-AUG-2003 to correct OS
 CC field.)
 XX
 SQ Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 7.03e-11 Length: 50937
 Score: 648.50 Matches: 684
 Percent Similarity: 29.82% Conservative: 286
 Best Local Similarity: 21.03% Mismatches: 1254
 Query Match: 4.91% Indels: 1041
 DB: 3 Gaps: 139

US-09-522-753-5 (1-2517) x AAA09469 (1-50937)

QY 13 ArgAlaThrGluProArgTyrProHisSerLeuSerTyrProValGlnIleAlaArg 32
 DB 17535 CGAGAACGCTCACCACGACCCACCGCATCAGCCATCAGCGCA-----17491
 QY 33 ThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp-TyrAlaSerHi 52
 DB 17490 -----CAGCACGCTCGGGATCCATCGCCGCA 17464
 QY 52 sLeuSerProGlySerIleIleGlnProGlnArg-----ArgArgProSerLeuLe 69
 DB 17463 CCCCCGACGGGCAACTC-----CCGCGCGGACACCCACCGACGCCATCCCTCAC 17410
 QY 69 uSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHi 89
 DB 17409 CAGCCCATATCCCCCAGCCACACATCTGTCGCGCGGACACCAAC---ACCCCGACGACGCA 17353
 QY 89 sSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProAr 109
 DB 17352 C-----CGCCAAACGCATCAAGAAAGCATTCGCGCGC 17320

QY 109 gLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGly-----127
 DB 17319 CATACACGCTGCCCCC-----ACTGCCCCACACACCGGCATTGGAGAGA 17272
 QY 128 -----GlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuTh 142
 DB 17271 ACAACACAAACGCTCCAGACACACAGGATCCACCAACTCACCCA-GAT-----17224
 QY 142 rGlyLysLeuGluProValSerProSerProPro---HisThrAspProGluLeuGl 161
 DB 17223 -----TCACCCGACCGCCACCTTGGCGCCCATCATACGCGCTCTCTGGACA 17174
 QY 161 uLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspAr 181
 DB 17173 GAGATCTCCGCCAAAGCGCTCGACTCAGGAAC--ACCAGCGCATGGACACCGCCGTCA 17116
 QY 181 gGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGl 201
 DB 17115 CCGGCTTCACCCAGATCCGACACAA-----CGCCCGACACGCCACGAT 17071
 QY 201 uGluGluAlaAlaLysPro-----ProGluProGluLysProValSer---ProPr 217
 DB 17070 CAGCCACATCAGAGCGCCGACCGACCGAGCCCGCCCAACCTTCAGTCTCCGCA 17011
 QY 217 oProIleGluSerLysHisArgSerLeuValGlnIleIleTyr-----231
 DB 17010 CCAGATCCCGACGACCGAGCGCACTGCCACACGCGCTCACAGAACCACCATGATCCG 16951
 QY 232 -----AspGluAsnArgLysLysAlaGluAl 240
 DB 16950 CCCCACACCCACCAACACCGCGCGTATGCGCACCCCAACCAAGACCCCGGTGA 16891
 QY 240 aAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPr 260
 DB 16890 CMAACACCGTCCACCGCGG-----CCGCCAACCCACCCACCATCCACCCAGGCAC 16837
 QY 260 oSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLe 280
 DB 16836 GCACACGACGACGACCCACAC-----ACCGAACCCACCGACCGCCACCT 16792
 QY 280 uIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGl 300
 DB 16791 GATCTCTAAACACGACGCGCACACACCCACCAACCGCGCGCACACCTCTCATCAA 16732
 QY 300 nArgTyrAspGlnLeu-----305
 DB 16731 CCCCACCGCGAGGTCCACCGACCCACCCCAACTCGGGNTGCTCCAGACCCAGAACAC 16672
 QY 306 -MetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAlaLy 325
 DB 16671 GCCCGAAACCCACACCTGCGCTCGACCGATCAATCACACACCGCATCCGAGGAC 16612
 QY 325 sGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGl 345
 DB 16611 CAGCAACCCACCGACCCCGCGTCCACGACCAACCGCGCGCTCTACCAACCCGCGCAA 16552
 QY 345 uLeuGlnGluArgMetGln-----SerArgValGlyGlnArgLysSerGlyLeuSerMe 363
 DB 16551 GATCAACACGCTCGGCCCAACACCGACGACCGCGGACAGCCGCGGGGAAACACAGAA 16492
 QY 363 tSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGl 383
 DB 16491 GCTC-----CGAGTCCCGCGCATCATCCAGGACAAGAACACGACAGA 16450
 QY 383 uAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMet-----398
 DB 16449 CCGCGCTCACACCGCGCGCGACAGCTCTCCGATAAGCCCGCGGTTCAGGACGGGTG 16390
 QY 399 -----LeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMe 416
 DB 16389 GATCGAGCTCCACGACACTGAC-----CT 16366

| | | | |
|-----------|-------|--|-------|
| Qy | 2162 | oLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgProProSerAs | 2182 |
| Db | 34155 | GCCTTCCCAACAGCGCGCGTGCACCGTTG-----CCGCCT----- | 34194 |
| Qy | 2182 | pLeuTyrProProAspHisGlyAlaProAlaArgGlySerProHisSerGluG1 | 2202 |
| Db | 34195 | -----ATACCTCCGATCGCGCGATCCCGCGCGCGCGCGTTCGGTAGACA---- | 34242 |
| Qy | 2202 | yGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyAspGlyI1 | 2222 |
| Db | 34243 | -----AATCCGCGGACCG-----CCGACGCC-ACCATTGGTCCGCGCGCGCGCGA-- | 34289 |
| Qy | 2222 | eGluProValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTy | 2242 |
| Db | 34290 | -----CCCGCGCGCGCGCGAACAACAGGCGTTCGCGCC----- | 34322 |
| Qy | 2242 | rProLeuLeuTyrArgAspGlyGluInThrGluProSerArgMetGlySerLysSerPr | 2262 |
| Db | 34323 | -----GGCACCACCGTTAGCGCGGTCCCGCGCGCGCGCGCGCC | 34363 |
| Qy | 2262 | oGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMe | 2282 |
| Db | 34364 | CGCGTTGCGCGTTCAACACCGCGC-----GGATCCGCGACACACCGCGCGCGCGCGG | 34417 |
| Qy | 2282 | tValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluPr | 2302 |
| Db | 34418 | -----CCCGCGCGGACCGCGGACCGCGGTTCGCGAACAACCGCGCGCGCGCGCGCCA | 34476 |
| Qy | 2302 | oGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh | 2322 |
| Db | 34477 | CCGACTTGACCGCGCGCGCGCGACCGCGGTACCGCCATTACCCCAACAACCGCGCG | 34536 |
| Qy | 2322 | rglyLeuMetThrTyrArgSerGlnAlaValGlnGlu-----HisAlaSerThrAsnMe | 2340 |
| Db | 34537 | CCCCCACCAGCGCTCCCGCGTCCCGCGCGCGCGTGAACCGCATCACCGATCAGCGGCGC | 34596 |
| Qy | 2340 | tGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlu----- | 2358 |
| Db | 34597 | CCCAACACACACTGTG-----TGCGCGCGTTG | 34623 |
| Qy | 2359 | -GluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPr | 2378 |
| Db | 34624 | ATCCACCCACACTGTGCT-----CC | 34647 |
| Qy | 2378 | oAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerPro-- | 2397 |
| Db | 34648 | AGCGCTGCAGCGGTGATTCGCGCGCTCCGCGCGTGCATACGCGCTCCAGCGCGCG | 34707 |
| Qy | 2398 | ----GlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSe | 2416 |
| Db | 34708 | GTACGCGAGCGCAAACTGCTCATGAACGTCGCCACCGCGCGCTCAACGCTGGTAC | 34767 |
| Qy | 2416 | rProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGl | 2436 |
| Db | 34768 | TCCTGCGGTGGGTACCAACAACACCGCGCATCGCGCGCGACACTCAT----- | 34816 |
| Qy | 2436 | uGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSe | 2456 |
| Db | 34817 | -----CACCGCGCGCGCGCA-----ACACTGCGTC | 34842 |
| Qy | 2456 | rAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMe | 2476 |
| Db | 34843 | GTGCGCGCGCGTGC-----CCGATTCCGCGCGGTGATGGCTGCCCAATCC-----CG | 34893 |
| Qy | 2476 | tAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaGly----- | 2493 |
| Db | 34894 | GTCAAGTCCGCGCGCGCGCGCCACAGCTCCGCGCGCCCATCAGCGACATGACCATT | 34953 |
| Qy | 2494 | -ProHisAlaTrpAspGluPro | 2502 |
| Db | 34954 | CCTCAACACCAATGCGGTACAGCG | 34981 |
| RESULT 46 | | | |

| | | | |
|------------|---|----|---|
| AAA09469/c | AAA09469 standard; DNA; 50937 BP. | | |
| ID | AAA09469 | XX | 06-AUG-2003 (revised) |
| AC | AAA09469; | XX | 29-AUG-2000 (first entry) |
| DT | 06-AUG-2003 (revised) | XX | Streptococcus oleandomycin gene cluster. |
| DT | 29-AUG-2000 (first entry) | XX | oleandomycin; oleandolide; polyketide synthase; oleAl; oleAlI; oleAlII; |
| XX | Streptococcus oleandomycin gene cluster. | XX | PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase; |
| XX | oleandomycin; oleandolide; polyketide synthase; oleAl; oleAlI; oleAlII; | XX | acyl-transferase; acyl carrier protein; inactivated; polyketide; |
| KW | PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase; | XX | macrolactone; antibiotic; motilide; erythromycin; ss. |
| KW | acyl-transferase; acyl carrier protein; inactivated; polyketide; | XX | Streptomyces antibioticus. |
| KW | macrolactone; antibiotic; motilide; erythromycin; ss. | XX | Key |
| XX | Streptomyces antibioticus. | XX | Location/Qualifiers |
| OS | Streptomyces antibioticus. | XX | 152..1426 |
| XX | Key | XX | /tag= a |
| XX | Location/Qualifiers | XX | /label= oleI |
| XX | 152..1426 | XX | complement(1528..2637) |
| XX | /tag= a | XX | /tag= b |
| XX | /label= oleI | XX | /label= oleN2 |
| XX | complement(1528..2637) | XX | complement(2658..4967) |
| XX | /tag= b | XX | /tag= c |
| XX | /label= oleN2 | XX | /label= oleR |
| XX | complement(2658..4967) | XX | 5772..18224 |
| XX | /tag= c | XX | /tag= d |
| XX | /label= oleR | XX | /label= ORF1 |
| XX | 5772..18224 | XX | /product= "8,8a-deoxyoleandolide_synthase_1" |
| XX | /tag= d | XX | 18267..29717 |
| XX | /label= ORF1 | XX | /tag= e |
| XX | /product= "8,8a-deoxyoleandolide_synthase_1" | XX | /label= ORF2 |
| XX | 18267..29717 | XX | /product= "8,8a-deoxyoleandolide_synthase_2" |
| XX | /tag= e | XX | 29787..40346 |
| XX | /label= ORF2 | XX | /tag= f |
| XX | /product= "8,8a-deoxyoleandolide_synthase_2" | XX | /label= ORF3 |
| XX | 29787..40346 | XX | /product= "8,8a-deoxyoleandolide_synthase_3" |
| XX | /tag= f | XX | 40625..41830 |
| XX | /label= ORF3 | XX | /tag= g |
| XX | /product= "8,8a-deoxyoleandolide_synthase_3" | XX | /label= oleP1 |
| XX | 40625..41830 | XX | 41878..43158 |
| XX | /tag= g | XX | /tag= h |
| XX | /label= oleP1 | XX | /label= oleG1 |
| XX | 41878..43158 | XX | 43163..44443 |
| XX | /tag= h | XX | /tag= i |
| XX | /label= oleG1 | XX | /label= oleG2 |
| XX | 43163..44443 | XX | 44433..45173 |
| XX | /tag= i | XX | /tag= j |
| XX | /label= oleG2 | XX | /label= oleM1 |
| XX | 44433..45173 | XX | 45251..46411 |
| XX | /tag= j | XX | /tag= k |
| XX | /label= oleM1 | XX | /label= oleY |
| XX | 45251..46411 | XX | 46491..47714 |
| XX | /tag= k | XX | /tag= l |
| XX | /label= oleY | XX | /label= oleP |
| XX | 46491..47714 | XX | complement(47808..49517) |
| XX | /tag= l | XX | /tag= m |
| XX | /label= oleP | XX | /label= oleB |
| XX | complement(47808..49517) | XX | WO200026349-A2. |
| XX | /tag= m | XX | 11-MAY-2000. |
| XX | /label= oleB | XX | 22-OCT-1999; 99WO-US024478. |
| XX | WO200026349-A2. | XX | 29-OCT-1998; 98US-0106100P. |
| XX | 11-MAY-2000. | XX | 16-FEB-1999; 99US-0120254P. |
| XX | 22-OCT-1999; 99WO-US024478. | XX | (KOSA-) KOSAN BIOSCIENCES INC. |
| XX | 29-OCT-1998; 98US-0106100P. | XX | Betlach MC, Shah SK, Mcdaniel R, Tang L; |
| XX | 16-FEB-1999; 99US-0120254P. | XX | |
| XX | (KOSA-) KOSAN BIOSCIENCES INC. | XX | |
| XX | Betlach MC, Shah SK, Mcdaniel R, Tang L; | XX | |

WP AA199683_25 2500001 2610000
 WP AA199683_26 2600001 2710000
 WP AA199683_27 2700001 2810000
 WP AA199683_28 2800001 2910000
 WP AA199683_29 2900001 3010000
 WP AA199683_30 3000001 3110000
 WP AA199683_31 3100001 3210000
 WP AA199683_32 3200001 3310000
 WP AA199683_33 3300001 3410000
 WP AA199683_34 3400001 3510000
 WP AA199683_35 3500001 3610000
 WP AA199683_36 3600001 3710000
 WP AA199683_37 3700001 3810000
 WP AA199683_38 3800001 3910000
 WP AA199683_39 3900001 4010000
 WP AA199683_40 4000001 4110000
 WP AA199683_41 4100001 4210000
 WP AA199683_42 4200001 4310000
 WP AA199683_43 4300001 4403765

Alignment Scores:
 pred. No.: 1.06e-10 Length: 110000
 Score: 651.00 Matches: 536
 Percent Similarity: 30.24% Conservative: 188
 Best Local Similarity: 22.39% Mismatches: 928
 Query Match: 4.93% Indels: 747
 DB: 4 Gaps: 95

US-09-522-753-5 (1-2517) x AA199683_37 (1-110000)

QY 579 LysGlyAlaThrArgSer---MetAlaAsnGluAlaAsnSerGluGluAlaIleThr 597
 DB 28635 AAACCCAGATTGACGGCGAGCGAGCTTGGCGGCTCGGCG----- 28673

QY 598 ProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsn----- 610
 DB 28674 -----GCGGCGCGCGAGCTTGCCAGCGGCGGCAACCGCGCTCGGCGCC 28727

QY 611 ---GluSerSerArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuGlu 629
 DB 28728 GCCGAGCGCGGTGTGTACCCAGCATCGCGGCCACGTCCTGGGCCACATCAGCTCG 28787

QY 630 HisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCys 649
 DB 28788 TAGTCGAACCTCCGCGCGCGGATCGCGGCTGTCTTGGCGCAACAGA----- 28835

QY 650 LysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnHis 669
 DB 28836 -----TTCGATAACGCCAGCGACACTAACCTCGACCGATTGGCGCGATGACG 28883

QY 670 Lys-----LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAla 686
 DB 28884 AAGGGTCCACCTCGCGGCGCAACCGCGCTCGACACACCCACCACCGCGGCGCTGC 28943

QY 687 ProAlaAlaSerGluGluAlaAla----- 695
 DB 28944 CCGGCGCGGAGCGCGCGCGCTCAACCAACCCCGCATACGGGCGGCGCC 29003

QY 696 -----PheProVal-----Val 700
 DB 29004 GCGCGCCCATCGCGACCGAGGCGCCCTTCGCAGATACCCAGCAGCCCGGAGGTC 29063

QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
 DB 29064 ACCGACCGGAAGCGCGCGCGC-----GAGCCACGCTCGGCGGCGCAGC 29108

QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
 DB 29109 TCATCCAGGCGCGCGCGCGCGCAACAGGGGGCC-----GGACCGCC 29153

QY 741 ThrValAsnAsnSerSerAsp-----ThrGluSerIleProSer 753
 DB 29154 CCGGTATATATACAGCGGAGTTGATCTCTGGGCGGATTCAGCAAAACATCATGCGCGCA 29213

QY 754 ProHisThrGluAla-----AlaLysAspThrGlyGlnAsnGlyProLysProPro 770
 DB 29214 GCCCTTTCGCGTTCGATCCCAACATCGTGTCAACCGGTGATCAGGCTGTGGCGCG--- 29270

QY 771 AlaThrLeuGlyAlaAspGlyProProGlyProProGlyProProProProProPro 790
 DB 29271 -----GCGCGCGAGCGCGCGCGCTCGCGCGCGCAACCTCGGCTCC 29309

QY 791 -----ArgAlaProIleGluProThrProAlaSerGluAlaThr 803
 DB 29310 GTGCGTGTGGTGGCGGCGCTCTTTCGCGCGCGCGCGCGCGCTTGGCGCG 29369

QY 804 GlyAlaProThrProProAlaProProSerProSerAlaPro----- 818
 DB 29370 CTGTTGCGCGCGTTCGCGCGCTCACCGCGCTCACCGCGCGAGCGCGTTCGCG 29429

QY 819 -----ProProValValProLysGluGluLysGlu 828
 DB 29430 CTCTGAGTGGCGCGCAATGCGCGCTGGCCACCTTACCGCGCTTGCACCAAGCCG 29489

QY 829 GluGluThrAlaAlaAlaProValGluGluGluGluGluGluGluGluGluGlu 848
 DB 29490 CCGTCCGGGCGGTTCGCTCGCA-----CGCGCGCG-GCC 29524

QY 849 GluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGlu 868
 DB 29525 GCCAAGCGCGCTTTCGCGCGGTGAGCC-----GCGCGCATTCGCGCGCTGCCACC 29578

QY 869 GluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaGluAla----- 885
 DB 29579 GAGCGCGCTTCGCGCGCGCACCGCGCAAGACGCGCTCGCGCGCGCGCGCACACC 29638

QY 885 ----- 885
 DB 29639 GCGTTGCGCGCGCGCACCGCGCGCGCGCTACCGCGCGCTTGGGCGCGCGCTG 29698

QY 886 -----ThrAlaGluGlyAlaLeuLysAlaGluLys-----LysGluGlyGlySer 900
 DB 29699 ACGGCGTTCGCGCGCGCGCTTTCGCGCGGAAAGCGCGCTGCGCGCGCGCGCG 29758

QY 901 GlyArgAlaThrThrAlaLysSer----- 909
 DB 29759 GCGCGCGCGCGCACCGCGCTTGGGTGAGCGAGCTTTCGCGCGCGCGCGCTC 29818

QY 910 -----GlyAlaProGln 913
 DB 29819 ACCGCGTGGCGCACCGTTCGCGCGCTTCGCGCGCTCGCGGAGCTAAGGCTGCGTTC 29878

QY 914 AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAsp 933
 DB 29879 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTTGGCGCG 29938

QY 934 LysAsnArg-LeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAl 953
 DB 29939 GATCAGCGCGCGCTTCGCGCGCGCGCGCGCGCTT-----CGCGCTT 29980

QY 953 aAsnAlaSerProGlnLysProLeu-----AspLeuLysGlnLe 966
 DB 29981 GCGCGCGTTCGCGCGCGCTTGGGTGTGTCGCGCGCGCGCGCGCTTGGCGCGTT 30040

QY 966 uLysGlnArgAlaAlaIlePro-----ProIleGlnValThrLysValHi 982
 DB 30041 GCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTGAGCGCGAGCTCC 30100

QY 982 sGluProProArgGluAspAlaAlaProThrLysProAlaProProProProPr 1002
 DB 30101 ATCACCAGCATTCGCCACCGCGCGCGCTTCGCGCGCTGAGCGCGTGTACCGCGCG 30160

QY 1002 oGlnAsnLeuGlnProGluSerAspAlaProGlnGln----- 1014
 DB 30161 GCCACGTTGCCCGCGGAGCGCGCGCGCGCGCGCGCTGCCGTGCCCGGTTGGTGGC 30220

```
Qy 2039 rGserLeuGlyTyrHisGlySerSerTyr---SerProGluGlyValGluProValSerP 2058
|||||
Db 4860 GGAGCAGCCGCGTCCGCGGTACCGGTCTCAGGACGACGCGGTCTGATGCTCTGACC 4801
|||||
Qy 2058 roVal-----SerSerProSerLeuThrH 2066
|||||
Db 4800 CACTCCGGCGCTCTCGATGATCAGATGCGCGTTCGTCCCACTGATCCCGAAGCAGCAC 4741
|||||
Qy 2066 isAspLys-----GlyLeuProLysHisLeuGluGluLeuA 2078
|||||
Db 4740 ACAGCCGCCGCGCGCGGTCCGACGCGCGGGCCACGCGCGCTCTGCGCAGCAGG 4681
|||||
Qy 2078 spLysSerHisLeuGluGlyGluLeuArgPro-----LysGlnProGlyProValLys- 2095
|||||
Db 4680 GC-----GACGGTCCCGACGACGACGTCGACGTGCGGTGCGCGCTCCGCG 4633
|||||
Qy 2096 -----LeuGlyGlyGluAlaAlaHisLeuProHisLeuA 2107
|||||
Db 4632 TGAAGTGTCCGCGCGCGGTCCGCTGCTCCA-GGGCAAGCAGCATCTTGACCAACCCGC 4574
|||||
Qy 2107 rGProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValL 2127
|||||
Db 4573 GCGCC-----CCGCGCGCGC 4559
|||||
Qy 2127 ysGlyHisGlnArgValValThrLeuAlaGlnHisLeuSerGluValIleThrGlnAspT 2147
|||||
Db 4558 CTGGGTGTGACCGACGTTGGACTTCAGCGACCCCGAGCC----- 4521
|||||
Qy 2147 yrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheP 2167
|||||
Db 4520 --ACAGGGTCCGCCA-----CCGCTCAGCCCGTACGTGGC 4484
|||||
Qy 2167 roGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProP 2187
|||||
Db 4483 CAGCAGCGCGTCCGCTCGATCGGAT---CGCCACAGCGCGCTCCCGTACCGTCCGCTC 4427
|||||
Qy 2187 roProAspHisGlyAlaProAlaArgGly-----S 2197
|||||
Db 4426 CACCGCGTCCAGTCAACCGCGCGCGAGCGCGGTCCGCCAGGGCAGCTCGATGACGCC 4367
|||||
Qy 2197 erProHisSerGluGlyGlyLysArgSer----- 2206
|||||
Db 4366 CTCTCTCGCGGAGCGGTTGGGAGCGGTGAGCGCGTGTGTCGACCGTCTCTGTTGACCGC 4307
|||||
Qy 2207 -----ProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspLyleGluProV 2225
|||||
Db 4306 ACTGCCCGCACCCAGCC----- 4290
|||||
Qy 2225 aliserProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuL 2245
|||||
Db 4289 -----CCAGCACCCGCTGTCCCGCACGCGCGGCGTCCGAGAGCCGCTCCAGTGCAGTAC 4235
|||||
Qy 2245 euTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnT 2265
|||||
Db 4234 G-CGCACACCTCCGACAGGCGGTACCG-----TCGGCATCGCGGAGAACG 4188
|||||
Qy 2265 hrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysS 2285
|||||
Db 4187 GCATCAGCGCGCGTGGGGCGAGCCCGCGTCCGCGGAGAACCTCCACG----- 4138
|||||
Qy 2285 erLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrA 2305
|||||
Db 4137 -----AACATGCCGGCGCCCGACATCA 4116
|||||
Qy 2305 snliserGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuM 2325
|||||
Db 4115 CCGTGGCGCGCGCGCACGACGACCGCCGCTACCTCGCCCGG-----CGCAGCGGCT 4065
|||||
Qy 2325 erThrTyrArgSerGlnAlaVal-----GlnGluHisAlaSerThrAsnMetGlyLeu- 2342
|||||
Db 4064 GCACGCGCAGGTGACGCGCCACCCAGGAGCAGGACCGCGGTGTCAGCGGTGAGCGCG 4005
|||||
Qy 2343 -----GluAlaIleIleArgLysAlaLeuM 2351
|||||
```

```
Db 4004 GTCCGCCGCGTCCGAGGGTGTAGGCGATCCGCCCGAGGCCACGCTCGCGGTGGTGC CGG 3945
|||||
Qy 2351 etGlyLysTyrAspGlnTrpGluLeuSerProProLeuSerAlaAsnAlaPheAsnProL 2371
|||||
Db 3944 TGAGCAGGTACCCCTCGTATCCGTCGCTCCCTCGTGAAGCCTCGGCCCGTACTCCGGTG 3885
|||||
Qy 2371 euAsnAlaSerAlaSerLeuProAlaAlaMetPro-----IleThrAlaA 2386
|||||
Db 3884 CGGTGGCGCGCAGCAACACTCCGGCGTCCGCCCGCGCAGTGACTCGGGAGCAGCACTCCGG 3825
|||||
Qy 2386 laaspGlyArgSerAspHis-ThrLeu-----ThrSerProGlyGly 2399
|||||
Db 3824 CGTCTCGAGGGCTCCACAGCGCTTCCAGCAGCAAGCGCTGCTCCGGGTCCATGGCGG 3765
|||||
Qy 2400 GlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro 2419
|||||
Db 3764 TGGCTTCGCGCGCGGATATCC---CGAAGAACTCCCGCTCGAACTCCGCCGCGC 3714
|||||
Qy 2420 GlyLeuAlaSerGlyAspArgProPro-----SerValSerSerValHisSerGlu 2436
|||||
Db 3713 -----AGTCAGGAACCCGCCCTCCCGCACATAGTCTTCCCGGCGCTCCCGGCTCG 3660
|||||
Qy 2437 GlyAspCysAsnArgThrProLeuThrAsnArgValTrpGluAspArgProSerSer 2456
|||||
Db 3659 GGTCTACAGCCCGTCCAGGTCCAG-CGCGCGTCTGTTGGGAGGACGAGATGGCGTGC 3601
|||||
Qy 2457 AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet 2476
|||||
Db 3600 GCG-----CCGACTCCAGCAGCTCCCGAGAGCGCGCGC 3565
|||||
Qy 2477 AlaSerProProPro-----ProGlyLeuProAlaGlySerGlyPro 2490
|||||
Db 3564 GAGTCGACGCGCGCGGAGCGGAGCGCAGCACCCACGATGGCGATCGGTCTCGTGC 3505
|||||
Qy 2491 LeuAlaGlyProHisHisAlaTrpAsp-----GluGluProLysPro--- 2504
|||||
Db 3504 TCCGCGGTGCGCGCGCGCTCCGACACCCCTACCGCGCGCCGACGAGACACTCCGGGC 3445
|||||
Qy 2505 -----LeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
|||||
Db 3444 CGCGCGCGCTCGAAGCTCTTCTCGCGCGGTGCGGAGCGCGCGGGGAG 3391
|||||
```

RESULT 45

AAI99683_37
Continuation (38 of 44) of AAI99683 from base 3700001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | AAI99683_00 | 1 | 110000 |
| WP | AAI99683_01 | 100001 | 210000 |
| WP | AAI99683_02 | 200001 | 310000 |
| WP | AAI99683_03 | 300001 | 410000 |
| WP | AAI99683_04 | 400001 | 510000 |
| WP | AAI99683_05 | 500001 | 610000 |
| WP | AAI99683_06 | 600001 | 710000 |
| WP | AAI99683_07 | 700001 | 810000 |
| WP | AAI99683_08 | 800001 | 910000 |
| WP | AAI99683_09 | 900001 | 1010000 |
| WP | AAI99683_10 | 1000001 | 1110000 |
| WP | AAI99683_11 | 1100001 | 1210000 |
| WP | AAI99683_12 | 1200001 | 1310000 |
| WP | AAI99683_13 | 1300001 | 1410000 |
| WP | AAI99683_14 | 1400001 | 1510000 |
| WP | AAI99683_15 | 1500001 | 1610000 |
| WP | AAI99683_16 | 1600001 | 1710000 |
| WP | AAI99683_17 | 1700001 | 1810000 |
| WP | AAI99683_18 | 1800001 | 1910000 |
| WP | AAI99683_19 | 1900001 | 2010000 |
| WP | AAI99683_20 | 2000001 | 2110000 |
| WP | AAI99683_21 | 2100001 | 2210000 |
| WP | AAI99683_22 | 2200001 | 2310000 |
| WP | AAI99683_23 | 2300001 | 2410000 |
| WP | AAI99683_24 | 2400001 | 2510000 |

| | | | |
|----|------|---|------|
| Qy | 1705 | GlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyr | 1720 |
| Db | 5663 | GCCATTCA | 5655 |
| Qy | 1725 | AlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuVal | 1744 |
| Db | 5654 | -----CGCGCGNACCGACACCCACT | 5634 |
| Qy | 1745 | ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAla | 1764 |
| Db | 5633 | CCCCCAACCCCTCAACAACCCCGCGACACCCAGTCCGACGCGCCCGACCGACCATCC | 5574 |
| Qy | 1765 | ProGlnProPheSerSerArgHisSerSerProLeuSerProGlyGlyProThrHis | 1784 |
| Db | 5573 | CCCCACCCACCGGCAACCCACCATCAACCCCGCGCCACCA-----CCAACCGCA | 5520 |
| Qy | 1785 | LeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgGluArg | 1804 |
| Db | 5519 | CCGCATCACCACGACGACACACCCCGCCACATACGACGACGCCACTCCCCACCGAAT | 5460 |
| Qy | 1805 | AspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrValGluHisAla | 1824 |
| Db | 5459 | GAC-----CCACACACCCGACACCTCCACACCCC | 5430 |
| Qy | 1825 | ProIleTrpArgProGlyThr-----GluGlnSerSer | 1835 |
| Db | 5429 | GAGCCTCCAAACGCGCCGGAACAACGCCACCTCCAAACGCAACACGCGAGGTGAGCAAACT | 5370 |
| Qy | 1836 | GlySerSerGlySerGlyGlyGlySerSerSerArgProAlaSerHisSer | 1855 |
| Db | 5369 | CGTCCGACCCACACCCACCCACGACGAGACCGGACACCGACCC---CAGCACCCG | 5313 |
| Qy | 1856 | HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgPro | 1875 |
| Db | 5312 | CACCGGACCGCATCATCGCCCGGACCCCGCTCCA-----CGTCGACGTGCG | 5265 |
| Qy | 1876 | SerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysPro | 1895 |
| Db | 5264 | CGAACATCATCATCCCGCA-----ACGACGACCATTCACC--TCCCCC | 5224 |
| Qy | 1896 | ThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProPro | 1915 |
| Db | 5223 | ACCATCGACAACACCTCATCCAGCACCCTCGCAACACCCACCCCGCATACAACCCA | 5164 |
| Qy | 1916 | -----AlaThrHisCys-----ProLeuGlyGlyThrLeuAspGlyValTyr | 1929 |
| Db | 5163 | CGCCCCCATCAACACCACTCGCACCCCTGACCCCGTGAAACAACACC----- | 5119 |
| Qy | 1930 | ProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGlu | 1949 |
| Db | 5118 | CCGACC-----CCACCCGCGACAGCCCGCCCGCGGAC | 5089 |
| Qy | 1950 | ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeu | 1969 |
| Db | 5088 | CGCACACCCCGACGACAC----- | 5071 |
| Qy | 1970 | GluProAlaSerSerProSerLysGly-----SerGluProArg-----Pro-LeuValPr | 1986 |
| Db | 5070 | CTCCCGGACGACAAACCCCAAGACGACGACGAGTCTCCCGACGCCACCAACGCGGACCC | 5011 |
| Qy | 1986 | oProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHi | 2006 |
| Db | 5010 | ACCCGACGGTGCTCGAAAGAGTGCCGCTCGAAGCCGCAATAAG-----CCACATCC | 4957 |
| Qy | 2006 | AlaSerProAspProPro-----AlaProProAlaSerAlaSe | 2019 |
| Db | 4956 | GGCGGATCCGACCCACCCGCTCCACGCTACTCACCCAAACCGACCGGCTCGCGCCGCAAC | 4897 |
| Qy | 2019 | rAspProHisArgGluLysThr--GlnSerLysProPheSerIleGlnGluLeuGluLeuA | 2039 |
| Db | 4896 | GCCCCCTCCGACCGCGGGAACAACACGCG-----CGAC | 4861 |

QY 809 -ProProAlaProProSerAlaProProValProValProLysGluGluGluGlu 828
 Db 8424 TTGCCCGGCACCCCGGCTCCGGTGTATACAGCCCTCGACGTCCTCCAGCCAGCCAGTCCGCC 8365
 QY 828 uGluGluThrAlaAlaAla-----ProProValGluGluGluGluGluGlu 844
 Db 8364 GGGAAACGGGAGAGCGTGTCCCGTATACCGCCACCAAGTCCACAGGTCTCCCGCGGAC 8305
 QY 844 sPro----- 845
 Db 8304 GCCACACCGCCGGGTAAACGGCAGCCATCCCCAGCATCCGATCCGCTCACGCTCGTCG 8245
 QY 846 -----ProAlaAlaGluGluLeuAlaValAlaAspThrGlyLysAlaGluGluProVally 863
 Db 8244 GCCATGACCGCGGTG-GGGGGCGCGCTCCGGTCCCGCTCCGCTCTCTCGGACCGAAGAG 8186
 QY 863 sSerGluCysThrGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 883
 Db 8185 TTCTGTCGTCAGGTGGAGGCGAGCGAGGGGGGTCCGGTGTCTGAGAGCA----- 8133
 QY 883 laGluAlaThrAlaGluGluAlaLeuLysAlaGluLysGluGluGluGluGluGluGluGluGlu 903
 Db 8132 -----GCTGTTCGCGAGTTCAGCGCGTGGCAGCGTTCAGCGGTTCAGCGAGTTCGAG 8078
 QY 903 laThrThrAlaLysSerGlyAlaProGlnAlaAspSerAspSerSerAlaThrCysSerA 923
 Db 8077 A-ACGCTGAGGAGTTCGAACCGCGGACCGAGCGAGCGAGTCCGGTTCGACCGCGTCGG 8019
 QY 923 laAsp-----GluValAspGluAla----- 929
 Db 8018 CCGAGCGTGGCGAGTATCCCGGGGCTCCGAGCGCAGACAGCGTGCAGCAGCAGCAGCAGCAG 7959
 QY 930 -----GluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuL 945
 Db 7958 GCGCTTCGCGCGGAGGCGGCGAGAGCAGCGCGCGCGCTCCGCGCGCTCCGCTCTCTGTC 7899
 QY 945 euThrProThrGlyAspPro----- 951
 Db 7898 CGCTCCCGCGGTGCGCGTCCGCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCT 7839
 QY 952 -----ArgAlaAsnAlaSerProGlnLysProLeuAspL 963
 Db 7838 CCGCGCATGCGCGCACCG 7779
 QY 963 euLysGlnLysGlnArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 983
 Db 7778 TACGGCGGTGACCAACGC-----CCGACGAAGTCTCCCGAGTCCAGTCCAGT 7734
 QY 983 luProProArgGluAspAlaAlaProThrLysProAlaProAlaProProProProG 1003
 Db 7733 CCGCGCAGCAGGAGGAGTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7675
 QY 1003 lnaAsnLeuGlnProGluSerAspAlaProGlnGlnProGly-----SerSerProArgG 1021
 Db 7674 -----GCGCCACCACTCGTCGAGGACGTCCGAGCGCGCGCT 7638
 QY 1021 lylSerArgSerProAlaProAlaAspLysGluAlaAlaAlaAlaAlaAlaAlaAlaAla 1039
 Db 7637 CCGGGTCCAGGCGCGTTCG 7581
 QY 1039 laGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProP 1059
 Db 7580 CCATTCGGCAGCCG 7521
 QY 1059 roArgGluValIleLysAlaSerProHieAlaProAspProSerAlaPheSerTrAlaP 1079
 Db 7520 GGGCGCGTCCG 7489
 QY 1079 roProGlyHisProLeuProLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 1099
 Db 7488 -----TTGGCGCGCGGTAGACAGCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 7449
 QY 1099 roProThrIleSerAsnProProLeuIleSerSerAlaLysHisProSerValLeuG 1119

Db 7448 GCCGACGATCGAGGAGAGCAGCAGGCGTCCGAGCGGAAGGTCCGCGGTCCAGCTCGT 7389
 QY 1119 luArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHis----- 1134
 Db 7388 CCAGGTGACGGCGCGCGGTGACCTTGTCCGACGATACGCG--GTCACCCCGGTCCGAGGTG 7330
 QY 1135 -----ValProTyrSerGluHisAlaL 1142
 Db 7329 AGCAGTTCGAGCACACCGCTCGTCGAGCACACTCGCGCGTGCAGTACGGCGACGAGGGG 7270
 QY 1142 ysAlaProValGly-----ProValThrMetGly----- 1151
 Db 7269 C-GCTCTCGGGAGGCGCGGAGGAGCGCGCAGTCCGTCGCGTCCGCGTCCGCGACGTCCGA 7211
 QY 1152 -----LeuProLeu-----ProMetAspProLysLysLysLysLysLysLysLysLys 1165
 Db 7210 GGCCTCGAGCGGCACTTCGGAACCGAGCGCGTTCAGCTCCGACGAGTTCGCCACATCC 7151
 QY 1165 exGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyProPro-GluSerLeu 1184
 Db 7150 GGGG-----CGCGCGGCG-----CGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 7115
 QY 1185 GlyValProThrAlaGlnAlaSerValLeuArgGlyThrAlaLeuGlySerValPro 1204
 Db 7114 -----ACG 7085
 QY 1205 GlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp-SerAlaIleTh 1224
 Db 7084 GCGACCTCGCACCCAGGCG-----ACCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7036
 QY 1224 rTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleTh 1244
 Db 7035 -----CGGAGGAGCACCGCGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7004
 QY 1244 rArgIleIleGly-----GluAspSerProSerArgLeuAspArgGlyArgGluAspSe 1262
 Db 7003 GTCCGAGGTCTCGTTCGTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6945
 QY 1262 rLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluG 1282
 Db 6944 TCGCAGGAGGCGCG-----GACGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6888
 QY 1282 yGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHi 1302
 Db 6887 CGGAGG-GCGGTACCG-----CGAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6855
 QY 1302 sGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaI 1322
 Db 6854 -CGCGAGCAGCG 6811
 QY 1322 eSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerPr 1342
 Db 6810 GCATCC-----CGGCGAGGTTCGCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6775
 QY 1342 oHisHisLeuLysGluGlnHisIleArgGlySerIleThrGlnGlyIleProArgSe 1362
 Db 6774 CGCCCCAGTTCGAGCG 6738
 QY 1362 rTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLysLysArgGluG 1382
 Db 6737 -----ACAGCTGAGCG 6712
 QY 1382 yThr----- 1393
 Db 6711 GTCTCGTCCG 6652
 QY 1384 -Pro 1392
 Db 6651 ACCGCTCG 6592
 QY 1393 -----AlaTyrIleThrGlnAlaLeuGlyPr 1404

Db 10079 GAGCAAACTCCGTCGACGACCAACACCCACCCACCCAGTCAGCAT-CCGCGAATC 10021
QY 208 -----ProGluProGluLysProValSerPro----- 216
Db 10020 ACATCCGGCAACGACGACCATCTCCCCACACCCACACACCTCATCCAGCAC 9961
QY 217 -----ProPro-IleGlu---SerLysHisArgSerLeuValGlnIleI 230
Db 9960 TCGCAAAACACCCACCCCGCATACACCCAGCCGCCCCATACCAACCCACTCGCACCC 9901
QY 230 leTyrAspGluAsnArgLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyP 250
Db 9900 TGACCCGTGAACAACACCCACCCACCCAGCCGCGACCCG----- 9860
QY 250 roGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHis----- 267
Db 9859 -----ACTCCCGCCAGCACAGACCGCGCCGCCCCGAGACACACCCAGGAGCC 9811
QY 268 -----GluAsnIleLysIleAsnGln-----AlaMetArgLysL 279
Db 9810 CCGCAGCCACCATCTCAGACCGCCAGCAACTCTCTCAGCGCCCGCGCCACACCCACC 9751
QY 279 yLeuIleLeuTyrPheLysArgAsnHisalaargLysGlnTyrLysGlnLysPheC 299
Db 9750 GCACGCTCACCGAACGACGACGACGACACCAACGA----- 9713
QY 299 yGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnA 319
Db 9712 -----CCCCGCCACATCCCGCGATCCGACCC 9685
QY 319 snProArgArgArgAlaLysGluSerLysValArgLysTyrGluLysGlnPheProG 339
Db 9684 ACCCGCTCCACACTCAGCAACCGCAC-----C 9655
QY 339 luIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgLys 359
Db 9654 GCCTGCGCCGCAACGCCCCCTCGACCGCGAAGAACAC----- 9614
QY 359 erGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleAspGlyL 379
Db 9613 -----CCAGGACACGACCGCTCACCGCTCGGCGCGCTC----- 9578
QY 379 euSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetL 399
Db 9577 -----AGGTTACGGCGCTCCCGCTCGCTCCCGGCC 9544
QY 399 euTyrAspAlaAspGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspP 419
Db 9543 TC----- 9542
QY 419 roMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluT 439
Db 9542 ----- 9542
QY 439 hrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuG 459
Db 9541 -----CGCCGCGCTTCTCTCAG-----GACCACATGGCGGTGGTCCGCTCACCCCG 9493
QY 459 luArgLys-ThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsn 478
Db 9492 AACCCGACACCGCGCGCCG-----CGCACCCGCTCCACCCGCGCCG 9451
QY 479 TyrLysSerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGln 498
Db 9450 GGCCACGACCGGTCTCGGTACAGAG-CTCCACCGCACCGAGCCCACTCCACCTCGA 9392
QY 499 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 518
Db 9391 CGACGGCGCATCCACATGACGCGTCCGCGCAACGACCATGCGCATCCCTGTCAC- 9335
QY 519 GluLysAspGluLysGluLysGluLysGluLysGluLysGluLysProGluVal 538
Db 9334 -----CACCTTGATCACACCAACCCACCGCAGCGCGCTGCGCATGCGCGATGTT 9284

QY 539 GluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsn 558
Db 9283 CGACTTCAACGA----- 9272
QY 559 AspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArg 578
Db 9271 CCGACGCCAACAGGATCGCCAGCCCGCTCCCGCGCTAGTGGCCAGCAGCAGCCCGC 9212
QY 579 LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrPro 598
Db 9211 CTCGATCGGATCACCGCGCGCTCCCGTACCGTGGCGCTCCACCC----- 9164
QY 599 GlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu 618
Db 9163 -----ATCCACGCTCGCGGCGCACCGCCCGCATCAGCCAGCGCTCCAGGATCAC 9113
QY 619 GluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAla 638
Db 9112 CCGCTG-----CTGCGCCGAGCACCACTCGGCGCGCTCAGCCCATGTCT 9071
QY 639 ArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLys 658
Db 9070 CGCACCGTC----- 9062
QY 659 ArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsn 678
Db 9061 -----CTGATTCAGCGCACTGCCCGCACCGCCCA-GCACCCGATGACCGTTCGCGC 9009
QY 679 AlaArgArgLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProPro 698
Db 9008 GCGCGTCG-----AAAGCCGCTCCACGACGACACCCACAC-----CCTCCG 8964
QY 699 ValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetVal 718
Db 8963 CCCACCGCTCCGCTCG----- 8946
QY 719 GluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGly 738
Db 8945 -----CACCC-TCGCGCAACGCTTGACCGCGCTGCGCGCGCAGC 8905
QY 739 ProAlaThrValAsnAsnSerSerAspThrGluSerIle-ProSerPro----- 754
Db 8904 CCGCGCTGCGCGAGAACTCCAGAACCGCTGCGCGTCCGCTACCGCTCACCCACCG 8845
QY 755 -----HisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPr 770
Db 8844 GCCAACGCCAGATCACACTCACCCCGCGCAACGACTGCACCGCCAGATGCAACGCCACC 8785
QY 770 oAlaThrLeuGlyAlaAspGlyProPro-----ProGlyProProThrPro----- 785
Db 8784 AACGACGACGACACCGCGGTGTCCACCGTCAGCGCGCGCCCTCCAAACCCCAACGCATAG 8725
QY 785 ----- 785
Db 8724 GCAACCCGACCGGAGATCACATGCCGAGGGTCCCGTGAGCACATGCGCGCGTAGGTG 8665
QY 785 ----- 785
Db 8664 TCGCGCGCTCCCGAGCTGGGTGCGTAGCTCTGCTGGGAGATGCGCTGAGACACCG 8605
QY 786 -----ProArgArgThrSerArgAlaProIleGluProThr-- 797
Db 8604 GTCCGGGTGCGCGCAGCGAGTCCGGAACGATGCGCGCGCTCCAGCGCTCCACGAC 8545
QY 798 -----ProAlaSerGluAlaThrGlyAlaProThrPro----- 808
Db 8544 GTCTCCAGCAGCAACCGCTGCTCGGGTCCATGCGCGTGGCTGCGCGCGCGATATCCCG 8485
QY 808 ----- 808
Db 8484 AAGAACTCCGCTCGAACTCGCGCGCGGAAACGACGAGNACCCGCTTCCCGTACATAGCTC 8425

Db 447 GTACTCGGGTGGAGCTCTCGCTCAGCGCGGAGGCGGCACGCCGGCGCTCCAGCGC 388
Qy 2416 -----SerProAlaProGlyLeuAlaSerGlyAspArgPro----- 2427
Db 387 CTCCACGGCGTCTCCAGCAGCAGCGCTCTGGGGTCCATCCCTGGCGCTCATGAGG 328
Qy 2428 -----ProSerValSerSerValHis 2435
Db 327 CGAGATCCCGAAGAGCCCGCTGAGACATCATCCCGAGGAATCACCCTCGCG 268
Qy 2435 erGluGlyAspCysAsnArgThrProLeuThrAsnArgValTyrGluAspArgProS 2455
Db 267 CAGTAGCTCTTGGCCAGACGT-----CCGATCGGGGTC 232
Qy 2455 erSerAlaGlySerThrProPheProTyrAsnProLeuLeuMetArgLeuGlnAlaGlyV 2475
Db 231 GTAAACAGAGACGACTCCC-----ATCGGAGCGGAA 199
Qy 2475 alMetAlaSerProPro-----ProProGly-----LeuProAlaGlySerG 2489
Db 198 GCCATCGATCGCGCTCTCCCTCCCGCAGGATACGCCAGTAGTCTTCCGGCGTCCGCCAC 139
Qy 2489 lyProLeuAlaGly-ProHisHisAlaTyrAspGluGlu 2501
Db 138 GCCCTTGGCAGCGGCAGCCCATCGATACGATCGCGAT 100

RESULT 44
AAT80415/C
ID AAT80415 standard; DNA; 13987 BP.
XX
AC AAT80415;
DT 02-MAR-1998 (first entry)
XX
DE Hybrid srmg/tylg ORF1.
XX
KW Tyllactone synthase gene cluster; tylg gene; multifunctional protein;
KW platenolide synthase gene cluster; platenolide production; srmg gene;
KW polyketide; tyllactone synthesis; antibiotic; tylostin; hybrid gene; ss.
XX
OS Streptomyces ambfaciens.
OS Streptomyces fradiae.
XX
FH Location/Qualifiers
FT 350..13987
FT /*tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes hybrid protein shown in AAW22611"
XX
EP791655-A2.
XX
PD 27-AUG-1997.
XX
PF 19-FEB-1997; 97EP-00301056.
XX
PR 22-FEB-1996; 96US-0012078P.
XX
PA (ELIL) LILLY & CO ELI.
XX
Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
XX
WPI; 1997-418046/39.
DR P-PSDB; AAW22611.
XX
PT DNA encoding Streptomyces fradiae tyllactone synthase domain - for
XX production of tylostin-related polyketide compounds.
PS Claim 22; Page 178-197; 220pp; English.
XX
CC This sequence represents a hybrid gene of the invention. This sequence
CC was created by replacing a EcoRI-ApaI fragment of srmg ORF1 with a EcoRI-
CC SfiI fragment from tylg ORF1. The position of the nucleotides from each
CC of the two genes is not given in the specification. The srmg gene (see

CC AAT80414) was isolated from Streptomyces ambfaciens, and encodes the
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenolide. Platenolide is the basic building block of the macrolide
CC antibiotic spiramycin. The tylg gene (see AAT80413) is the tyllactone
CC synthase gene cluster of the invention. The tylg sequence was isolated
CC from Streptomyces fradiae, and encodes multifunctional proteins which
CC direct the synthesis of the polyketide tyllactone. Tyllactone is the basic
CC building block of the antibiotic tylostin. The hybrid sequence can be used
CC to transform S. ambfaciens lacking the srmg ORF1 sequence, or S. fradiae
CC lacking the tylg ORF1 sequence, so that they can produce polyketides. The
CC DNA sequence can be modified so as to alter the type of carboxylic acids
CC incorporated, the number of carboxylic acids incorporated and/or the post
CC -condensation reactions performed, thereby resulting in novel tylostin-
CC related polyketides
XX
SQ Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.03e-11 Length: 13987
Score: 652.00 Matches: 648
Percent Similarity: 30.21% Conservative: 260
Best Local Similarity: 21.56% Mismatches: 1034
Query Match: 4.93% Indels: 1081
DB: 2 Gaps: 133
US-09-522-753-5 (1-2517) x AAT80415 (1-13987)
Qy 15 ThrGluProArgTyrProProHis-----SerLeuSer 25
Db 10684 ACCGACACCGACACCCACCCATCCCGGAACCGACCGCTCCCGGCATGAGC 10625
Qy 26 TyrProVal-----GlnIleAlaArgThrHisThrAspValGlyLeuLeuGlu 41
Db 10624 CACCCAAATACCCCGGATCCCACTCCCGGAAACCCACCCACCGACACACCGCA 10565
Qy 42 TyrGlnHisSerArgAspTyrAlaSerHisLeu-----SerProGlySer 57
Db 10564 CACCACACGACACCGCGCGGACCGACCGCACTCCAAACGACTCCACAAACCCCGGAA 10505
Qy 58 IleIleGlnProGlnArg-----ArgProSerLeuLeuSerGluPheGln 73
Db 10504 CTC-----CCCAACACCGGCTCCATCAACACCGAATGAACCCATCGCACATCCAA 10451
Qy 74 Pro-----GlyAsnGluArgSerGlnGluLeuHis 83
Db 10450 CCGCGGCACTCCACCCATCCCGCATCAGCGTGACACACCGACTCCAGCACACCCAC 10391
Qy 84 LeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIle 103
Db 10390 ATCACCAGACACACACCGACCGCGGCGC----- 10361
Qy 104 GluSerLysArgProArgLeuGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 123
Db 10360 ATTACCGCGCGCAACCGACACCGACTCCCG-CCAAACCCCTCAACACCCCGCACCGC 10302
Qy 124 LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysArgSerGluLeuThrGly 143
Db 10301 ACTCCGAGCGCGCCACCGACCATCTCCCGACCGGCAACCCCGCAACCCATCAACC--- 10245
Qy 144 LysLeuGluProValSerProProSer-----ProPro 154
Db 10244 -----CACCCCGCGCACCAACCAACCGGACCGCATCACCACCAACACACCCCGCCA 10191
Qy 155 HisThrAsp-----ProGluLeuLeuValProArgLeuSerLysG1 170
Db 10190 CATACGAGCAGCGACCTCCCGACCGAATGACCAACACCGACACCTCCACACCCC 10131
Qy 170 uGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnIle 190
Db 10130 GAGCTTC---CAACGCGCGGAAACAA-----CGCCACTCAACGCAACACGCGGCT 10080
Qy 190 eSerLysLeuLysLysGlnGlnGlnGluGluGluGluAlaLysPro----- 207

1786 QY -----ThrLysProThrThrThrSers 1793
2235 Db TGGCGTCAGCTCCACCATCAGCCGCTCCACGCTCCACGACACACCGTCGA 2176
1793 QY erSerGluArgGluArgAspArgAspArgGluArgGluLysS 1813
2175 Db GCTCG----- 2171
1813 QY erileLeuThrSerThrThrThrValGluHisAlaProIleTTPArgProGlyThrGluG 1833
2170 Db -----AGCTGTTACCGCCCGTATCGCAAGGCT-----CCCGTACGGCGC 2128
1833 QY lnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaS 1853
2127 Db GATGTGCTCTCCACCTCCGACA-----CCGGCGCTCGACACGACAT 2086
1853 QY erHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnG 1873
2085 Db CA-----TCGCCCCCATCC----- 2072
1873 QY lnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProS 1893
2071 Db -----CCGACTGTGCGCAGCCGCTGCTGCGCAGCGCCACTACCGCGCTC---CCTC 2020
1893 QY erLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrP 1913
2019 Db CGCAAGCGACAGCGCTCCGACACACCGCGCGCGACACCT---CGCCCTGGTGTGGC- 1964
1913 QY heProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuM 1933
1963 Db --CCACCACCGCGCAGGCTCCAGCCCGCCAGCGAC----- 1931
1933 QY etGluProValLeuLeuProLysGluAlaProArgValAlaArg-----P 1948
1930 Db -----GCCACCGCGCGCCAGCAGCCAGCAGCATCGGACAGCGCGGCTGCAC 1882
1948 QY roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
1881 Db CAGTCCACCGCTCCAGCAGCGCTGCTCTCACTGCTCACCAGCGCTCGCGCGCAG 1822
1968 QY lyLeuGluProAlaSerSerProSerLysGlySerGluProArg-----ProLeuValP 1986
1821 Db CACCGACAGCAGACAGCCGCTCCACCGCGCGCGCGCTCGTGGGATCGCTCGACCGC 1762
1986 QY roProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaPro---- 2004
1761 Db CTCCG-----CGAACGCTGGCTCTGCTCCAGCA-----GCCCGCGGCC 1723
2005 QY -----HisHisA 2007
1722 Db CATCCCGCGCACTGGCTCCTTGGCGGGGAACAGAACACACACCTTGCTCGCGCAGC 1663
2007 QY laSerProAspProProAlaProPro-----AlaSerAlaSerAsp 2021
1662 Db CG---CCGTACCGCGCGACACCGCGCGGTGGCGCGACCTGGACAGCGCGCGACAC 1606
2021 QY roHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuLeuArgSerL 2041
1605 Db CTCCACCGCTCGGACAGCTCGCGCTCGCAGCGATGCG-----GAGAGCGGAAGTG 1552
2041 QY euGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro-----ValserP 2058
1551 Db CGTCCGCTGACGCGCTCTGTGGCAGCAGCTCGACACCCACCTCCGCGTCTCTTC 1492
2058 QY roValSerSerProSerLeu-----Thr- 2065
1491 Db CAGCACTTCGCCAGCGCTCCGCTGCGCGCCACCGAGGCTCTGCTCGACCCCGACAG 1432
2066 QY HisAspLys-----GlyLeuProLysHisLeuGluLeuAsp 2078
1431 Db CACAGCGGCAATAGCCAGCGCGCTCGCTCGAGGCTCGCGCGCGCGCTCGG 1372
2079 QY LysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGly 2098

1371 Db CGCGCGCTTCTCGAGGATGATATGCGCTTCTCGTCCGCTGATCCC----- 1327
2099 QY GluAlaAlaHisLeu-----ProHisLeuArgProLeu 2109
1326 Db GAACGACGACACCGCGCAGCAGCGCGCGCTGCGCGCGCAGCGCGCTCTTG 1267
2110 QY ProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyVal----- 2126
1266 Db CAGCAGCGACAGCGCTTCTTCCCA-----CCCAATGTGGCGGTCTCG 1222
2127 QY -----LysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGlu 2141
1221 Db CTGCTCCGATCAGCGTTCGCGCAGCAGCTCTGCTCGCAGCGCGACAC----- 1171
2142 QY ValIleThrGlnAspTyrThrArgHisProGlnGlnLeuSerAlaProLeuProAla 2161
1170 Db -----CATCTTGATCAGCGCGCCA-CACCGCGCGCG-----CCTGGCATGTCCAGGT 1121
2162 QY ProLeuTyrSerPheProGlyAlaSer----- 2170
1120 Db TCGACTTCGACGACCCAGGTACAGCGGAGCTTCGGGCTACGCTTCGGGCGGAACACCT 1061
2171 QY CysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHis 2190
1060 Db CGCCACAGCGCTC-----CGGCTCGATCGGCTCTCGAGGCTCGTCCGCTGCCAT 1010
2191 QY GlyAlaProAlaArgGlySerProHisSerGluGlyLysArgSerProGluProAsn 2210
1009 Db GCGCTCCACCGCTCGATGCTCTCGGCGCAGACCGCG-----ACGACGACGCGCTGCC 953
2211 QY LysThrSerValLeuGlyGly-----GluAspGlyIleGlu---Pro 2224
952 Db GGATCACCGCTGCTCGCGAGCGCGTTCGGCGCGTACAGCTCGCTCGACCGCTCT 893
2225 QY ValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeu 2244
892 Db GGTTCACCGCAGCAGCGATCACCCCGCAGCAGCGTTCGCGT----- 848
2245 QY LeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPro---Gly 2263
847 Db -----CGCGCGCGCTCAGCAGCGCTTCAGCAACAGCATCCACACCTTCGCGCC 794
2264 QY AsnThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVal 2283
793 Db AGCCCGCGCGCTCGCGCGCGAGAGCTCTTGACCGACCG-TCTCGGCGCATCCCC 735
2284 QY LysSerLysLysGlnGluLeuLeuLysLeuAsnThrHisAsnArgAsnGluProGlu 2303
734 Db TTG-----AGCGCGCTGAACCTCCACGAACAGCAGCGGGGTGCTC 696
2304 QY TyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGly 2323
695 Db ATCAGCTCACCCCGCGCGCAG-----GCCAG-GTCGATTCGCC 655
2324 QY LeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGlu 2343
654 Db CTGGCGCAGCGCTCGACCGAGGTGACGACACACAGCGA----- 613
2344 QY AlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTTPGluGlu-SerProProLe 2363
612 Db -----CGACGACGCGCGCTGCTCCACCGTGTGCTGGCGCTCGAAGCGCC----- 567
2363 QY uSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProI 2383
566 Db -ACAGCTAAGCCACCCCGCTGAAAGCAGCTCGCGCGCTCCCGCTCCCGTCCCTCGGTACC 508
2383 QY eThrAlaAla-AspGly-----ArgSerAspHisThrLeuThrSerProGlyGlyG 2401
507 Db GTCCAGCGCGCGAGGTCTATTCGCGAGAGAGCAGCAGTGTAGTCGAGGCCCATCGAGCCAG 448
2401 QY lyLysAlaLysValSerGlyArgProSerSerSerSerSerSerSerSerSerSerSer 2415

PT enzymes comprises all or a portion of one or more genes in the *tmbA* gene cluster of *Sorangium cellulosum*.

PT cluster of Sorangium cellulosum.

PS Disclosure: SEO ID NO 17; 73pp; English.

The invention describes a recombinant DNA vector that comprises all or part of one or more genes in the *tmbA* gene cluster. Also disclosed is a host cell that comprises the *tmbA* genes of *Sorangium cellulosum* or fragments of those genes. The DNA vector is useful in preparing hybrid polypeptide synthase (PKS) enzymes and the polyketides produced by such hybrid enzymes. The gene products of the *tmbA* cluster can be used to synthesize the polyketide tobramycin. This sequence represents the *Sorangium cellulosum tmbA* gene cluster under reading frame.

XX
SQ Sequence 11358 BP; 1488 A; 3388 C; 4691 G; 1791 T; 0 U; 0 Other;

Alignment Scores:

| | | | |
|------------|----------|----------|-------|
| Pred. No.: | 1.73e-11 | Length: | 11358 |
| Score: | 652.00 | Matches: | 659 |

Percent Similarity: 31.86%
Best Local Similarity: 22.58%
Conservative: 271
Mismatches: 1075

```

Query Match: 4.93% Indels: 926
DB: 9 Gaps: 139

```

US-09-522-753-5 (1-2517) x ADC26983 (1-11358)

Qy 17 ProArgTyrProProHisSerLeuSerTyrProValcIleAlaAArgThrHisThrAsp 36

D_b 7386 CCCCGACTGTGCGGCACCGCTGGCT-----GCGCAGGCGCACTACCCG 73

QY 37 valGlyLeuLeuGluTyrGln-----HisHisSerArgAspTyr----- 49

Db 7341 CGCTCCCTCGGTAAGCGACAGCGCTCCGGACACACCACGCGCGCGGACACCTCGCCCTGGCT 72

Qy 50 ---AlaSerHisIeuSerProGlySerIleIleGlnProGlnArgArgProSerLeu 68

Db 7281 GTGGCCCAACACCGCGCAGG-----GTCCAGCCCCAG-----CGACCGCACGC 72

Qy 69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisIeuArgPro---Glu 87

Db 7236 CGCGGCCAGACCCACGACATCGCGAACAGCGGGGTGCACCACTCCACCGCTCCAG 71

Qy 88 SerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArg 10

Db 7176 CGACGGTGCTCTCTCCCGGCCATCTCCGGCAGCACCGACAGCAC- - - AGACGAGCCCGT 71

Qy 108 ProArg-----LeuGluLeuLeuProAspPro----- 11

Dbb 7119 CCACGGCGGAGCGGCTCATCGACGCTGCA-CCGCCTCCGCGAACGCTGGCTCTGCT 70

Qy 117 -----LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAsp 13

Db 7060 CCAGCAGCGTGGGCCATCCCCG-----GCCATGGCTCCCTGGCCGGGGA-----70

QY 135 LeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProSerProPro 15

Db 7012 -----ACACGAAACACCACTTGCCTCGCGCACGGCCCGTGCCTCCCGCACACCGCCC 69

QY 155 ---HisThrAspPro---GluLeuGluLeuValProPro-ArgLeuSerLysGluGluLe 17

Db 6961 GGTGGCGCGACCTTCCACAGCGCCCGCAGACCTCCACCGCCTCGGACACGCTCGCG 6959

QY 172 u l l e G l n a s n M e t A s p A r g V a l A s p A r g G l u I l l e T h r M e t V a l G l u G l n G l n I l l e S e r L y 193

Db 6901 CTTCACCGATCGCGAGA-----GGCGAAGTGCCTCCGGTG-----698

QY 192 sLeuLysLysLysGlnGlnGlnLeuGluAlaAlaLysProProGlu-----200

Db
6864 -----CAGCGTGCGTGCGACCAACGTCGACACACCCACCTCGCGT 6865

QY 210 -ProGluLysProValSerProProPheLeuGluSerLysHisArgSerLeuValGlnIle 220

db 6820 G C C C T T C C A G C C A C T T G C C C C C A C G C C C G G C C T G C A T T C A C C G C A G C C T C G T C G G A C 6

```

Db 18093 GCACCGGCTCGTCTCGGGTGGCCCTCACCGTCCGCTCGTCCAGGCGCTCGGGACG 18152
QY 2080 -----
Db 18153 CCGGTATCGACACCCCGCTGTGGGCGCTGACCCGCGGCGCGTCTCCACCGGCGGTCCG 18212
QY 2081 ---HisLeuGluGlyGluLeuArgProLysGlnPro-----GlyProValLys--- 2095
Db 18213 ACAAGGTACCGCGCGCGTCCAGGCCACGAGTACCGGATCGGCTGGACCGCGCGCTGG 18272
QY 2096 -----LeuGlyGlyGluAlaHisLeuPro---HisLeuArgProLeu--- 2109
Db 18273 AGTCCCGGACGCTGGGCGGTGTGTGTCAGCTGCCGACACACGACGCGCGCGG 18332
QY 2110 -----ProGlu 2111
Db 18333 GCCAGGCGTCCGCGCGTCTCGCGCGCGCTCGCGGACGACGACGATCGCCCTGC 18392
QY 2112 SerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
Db 18393 GCTCTCCGCGCTTCCACCGCGGTATCGTCCGCGCG----- 18431
QY 2132 valValThr-LeuAlaGlnHisSerGluValIleThrGlnAspTyrThrArgHisH 2151
Db 18432 -----ACGCGCGCGCGCGTCCGCGCGCGACTG-----GAAGCCGCGCGCACCA 18479
QY 2151 sProGlnGlnLeuSerAlaProLeuProAlaPro----- 2162
Db 18480 CCTGTGTACCG-GCGGCTCCGCGACCTCGCCCGCGACCTGGCGCGTGTGTCGCGGAG 18538
QY 2163 -----LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgAr 2178
Db 18539 CAGGGCGCGAACACCTGGTCTGGTCAGCGCGCGCGCGCG-----AGGCG 18586
QY 2178 gProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPr 2198
Db 18587 CCGGCGCGCGCGAACTCCGCGCGCGAGCTGGCGGACCGCGCACCGGACGACCTCGCC 18646
QY 2198 oHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal-LeuGlyGlyG 2218
Db 18647 GCTGTGCACATACCGACCGTACCGCTCGCGCGCGCTCGCGCGCGCTCGAAGCGCGAG 18706
QY 2218 lyGluAspGlyIleGluProValSerProProGlyMetThrGluProGlyHisSerA 2238
Db 18707 GCGCG-----CACGCTCCGCGACGCTGGTGTGCACACCGCGCGCACCATC 18748
QY 2238 rgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetG 2258
Db 18749 GAACGTCA-----CACCTCGACGCCA-C 18771
QY 2258 lySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluS 2278
Db 18772 CACCTCGACACTTGCACCGCTCTCGCGCGCAAGGTACCGCGCGCGATCTCTGA 18831
QY 2278 erAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisA 2298
Db 18832 CGAACTCTCGACGACGAGGAGC-----TGGACGACTTCGT 18867
QY 2298 snArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA 2318
Db 18868 CCTGTACTCTCCA-----CCGC 18885
QY 2318 lalleThrGlyThrGlyLeuMetThrTyrArg---SerGlnAlaValGlnGluHisAlas 2337
Db 18886 CGGCATGTGGGCGAGCGCGCGCGCGCGGTATGTGCGGGGCAAGC----- 18932
QY 2337 erThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnT 2357
Db 18933 -----CCTATCT 18939
QY 2357 rpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerL 2377

```

```

Db 18940 GCGCGCCCTCGCGC-----AGCACCGCGC 18963
QY 2377 euProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerP 2397
Db 18964 CCGCCGCGCGCTGACCGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19023
QY 2397 roGly-----GlyGlyGlyLysAlaLysValSerGlyArgProSerS 2411
Db 19024 CGGCCGAGTCGACCCGACAGATGATCCGCGCAGCGCTGGAGTTCATGACCCGCGCT 19083
QY 2411 erArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValS 2431
Db 19084 CCGCTGACGCGCTCAACGCGGCGCTCGACGACGACG-----AGCAGGT 19128
QY 2431 erSerValHisSerGluGlyAspCysAsnArgThrProLeuThrAsnArgValTyrG 2451
Db 19129 CATCGCGTCCGCGAGCTGACTGGGAGAGTACACCCCGTCTACCTCGCGCGCGCC 19188
QY 2451 luAspArg-----ProSerSerAlaGlySerThrProPheThrProTyrAsnP 2466
Db 19189 CACCCGCGCTGTCGACGAGGTCCCGGAGGTGACGCGCTCACCGCGCTCGCGAGCAGA- 19247
QY 2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProA 2486
Db 19248 -----GCGCGGGGACCCCGC 19263
QY 2486 laGlySerGlyProLeuAlaGlyProHisHisAlaTyr----- 2498
Db 19264 CCGGGCG-----AGTTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19314
QY 2499 -----AspGluGluProLysProLeuLeuCysSer 2508
Db 19315 CGCAAGCTGCTGAGACCGTCCGACCGAGCGCGCTCGCTGCTGCTGCTGCTGCTGCTG 19361

```

RESULT 43

ADC26983/C
ID ADC26983 standard; DNA; 11358 BP.

XX ADC26983;

XX 18-DEC-2003 (first entry)

DE Sorangium cellulosum tmbA gene cluster tmbC DNA.

XX recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;
KW polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;
KW ds; tmbC.

XX Polyangium cellulosum.

XX Key Location/Qualifiers
CDS 1..11358
FT /tag= a
FT /product= "tmbC"

XX US2003054547-A1.

XX 20-MAR-2003.

XX 28-AUG-2001; 2001US-00942025.

XX 22-JAN-1998; 98US-00010809.

XX 31-AUG-1998; 98US-00144085.

XX 15-FEB-2001; 2001US-0271245P.

XX (JULI/) JULIEN B.

XX Julien B;

XX WPI; 2003-567100/53.

XX P-PSDB; ADC26984.

XX New recombinant DNA vector for preparing hybrid polyketide synthase

| | | | |
|----|-------|--|-------|
| Qy | 1474 | -----SerSerProLeuSerProGlyGlyPro-----Thr | 1783 |
| Db | 17112 | AACTGCACGCACACGCTCACCGCGAGGAGATCCGCGCGCGCGGTCGCGGTGAGTACG | 17171 |
| Qy | 1784 | HisLeuThrIlyPro-----ThrThrThrSerSerSerGluArgGluArg | 1798 |
| Db | 17172 | CCTCGCAGCTCCCCCATATCGAGGACCTGCACACAGAGATCTCTGACTGTCGCGGAG | 17231 |
| Qy | 1799 | AspArgAspArgGluArgAspArgGluArgGluArgGluArgGluArgGluArgGluArg | 1818 |
| Db | 17232 | TCGCGC-----CGCGAGCTTCGAGATCCCGTCTCTCTCGACCGTCCACG | 17276 |
| Qy | 1819 | ThrThrValGluHisAlaProIleThrArgProGlyThrGluGlnSerSerGlySerSer | 1838 |
| Db | 17277 | GCGACTGGCTCGACACACCGCTCATGCGCGCGGTACT----- | 17315 |
| Qy | 1839 | GlySerSerGlyGlyGlySerSerSerArg----- | 1850 |
| Db | 17316 | GGTACCGACCTTCGCGCGCGCGGTGCTGTTTCGCGGACGAGTCCGTGACCTGATCGCG | 17375 |
| Qy | 1851 | -----ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg | 1866 |
| Db | 17376 | CCGACCACTGCGCTTCATCGAGTCACTCCACCCCGTCTCGGCATGTCGTCGTCAG | 17435 |
| Qy | 1867 | ThrGlnAspAlaLeuGlnGln-----ArgProSerValLeuHisAsnThrGly | 1882 |
| Db | 17436 | ACATGATCGAGCGCGGAGTGGCGGTGTCCTCCGCGCACCTCGCGCGGACAAACG | 17495 |
| Qy | 1883 | MetLysGlyIleIleThrAlaValProSerLys-----ProThrValLeuArgSer | 1900 |
| Db | 17496 | CGGTCTCGACCGCTTCGTCGCGCGCGGAGTGTTCGTCGCGGTGTCGAGGTGCG | 17555 |
| Qy | 1901 | ThrSer-----ThrSerSerProValArgProAlaAlaThrPheProAlaThr | 1917 |
| Db | 17556 | ACTGGCGCGCGTTCGAGGGACCGTCTCGTCCCGGTTCGCGGTTCGCTCCCACT | 17609 |
| Qy | 1918 | HisCysProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeu | 1937 |
| Db | 17610 | --ATGCTTCCACGACAGAACCTGTGGCCATGGCGCGCGCGCGGAGCGCG-- | 17660 |
| Qy | 1938 | LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis | 1957 |
| Db | 17661 | TCACCGCGCGACCCCGAGACCGCCCGT----- | 17690 |
| Qy | 1958 | AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro-----AlaSer | 1973 |
| Db | 17691 | -----TCTGACCGCGCTCGAGGACGAGGAGCTCTCGCGCTCACCGCGCGCTCG | 17741 |
| Qy | 1974 | SerProSerLysGlySerGluProArg-----ProLeuValProProValSerGlyHis | 1991 |
| Db | 17742 | GCACGACGAGGACTCCGCTCGCGCGGTGTCGCGGTGCTCTCTCTGCGCGCGGCCC | 17801 |
| Qy | 1992 | AlaThrIleAla-----ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro | 2009 |
| Db | 17802 | GCAAGGAGGCTCCACCGTCCACTCTCTGGCGTACCGCGGACCTGGAAGCGCGTCA | 17861 |
| Qy | 2010 | AspProProAlaProProAlaSerAla-----SerAspProHisArgGluLys | 2025 |
| Db | 17862 | AGCTCCCGACGACGACCCCTGGACGCGACCTGGCTGCTGTCGCGCAGACGCGCTC | 17921 |
| Qy | 2026 | ThrGlnSerLysProPheSerIleGlnGluLeuGluArgSerLeuGlyTyHisGly | 2045 |
| Db | 17922 | ACACGAGCTCGCGGAGCGGTGGAGACGCGCGCGGCGCGGCTCGCGC----- | 17972 |
| Qy | 2046 | SerSerTyr----- | 2048 |
| Db | 17973 | TCGTCTCGACGAGAGTGCACCGACCGCGCGGTGCTGCGGAGCGCTCACCGACGCG | 18032 |
| Qy | 2049 | -----SerProGlyValGluProValSerProValSerSerPro | 2062 |
| Db | 18033 | ACGGCTGACCGGATCGTCTCCGCTCTCGCGCGCGCGGAGCGGACCGGTCCGTC | 18092 |
| Qy | 2063 | SerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSer----- | 2080 |

Db 14077 CGTCTCGACGAGTGTCCGCCAGCAGCCCTCACCGGTGTGGTGCACACCACCGGCGT 14136
QY 997 -----ProAlaProProP 1002
Db 14137 CGGCCACTACGGCCCGCTGACGGCTGACCGCGCGAGTTCCCGCGCTCACCGCGC 14196
QY 1002 roGln-----AsnLeuGlnProG 1008
Db 14197 CAAGTGGCCGGCGCGCCACCTCGACAACTCTCTGGCGACCGGCACTGACTTCT 14256
QY 1008 luSerAspAlaProGlnGlnProGlySer 1018
Db 14257 CATCTGTCTGGTCCATCCCGCGCTCTGGGAGCGCGACACAGAGCCCTACGGCGC 14316
QY 1019 -----ProArgGlyLysSerArgSerProAlaP 1028
Db 14317 TGGCAAGCCCTACTCTGACGCCCTCGCCCTGGCGCGCGCGCGCTCGCGCGAC 14376
QY 1028 roProAlaAspLysGluAlaPheAlaGluAlaGlnLysLeuProGlyAspProP 1048
Db 14377 CTCCA-----TCGCCTGGGTCCCTGGGCGCGCACCGG 14409
QY 1048 ystPthrSerGlyLeuProPheProValPro 1058
Db 14410 CATGGCGCGCAGACGCCCTCTCCGCACCTCTCGACGCCAGCGGCTCGGCCCTGCTCGA 14469
QY 1059 --ProArgGluValIleLysAlaSerProHis 1069
Db 14470 CCCGCGC-----CCGCCCTCACGAGATGCGCGCGCGGTGCGCCAGACGT 14520
QY 1069 laProAspProSerAlaPheSer 1082
Db 14521 CACGCTCACGCTCGCGAGCTGACTGACCGCGCTACGCGCGCTGTTCACTCTCGCGCG 14580
QY 1082 isProLeuProLeuGlyLeuHis 1099
Db 14581 GCCAGCGCGCTCATCAGCGACCTGCCCGAGGTCCGCGCTCGCGCGCGAGAACCC 14640
QY 1099 roProThrIleSerAsnProProLeuIleSerAlaLysHisProSerValLeuG 1119
Db 14641 CGCGACACCGCGGACGCTCCGAG-ATCGTCAGCGGGTCCGCTCGCTCGAGCGCG 14699
QY 1119 luArg----- 1120
Db 14700 AGCAACTGCGCTGTCCACGACCTGTGTGCGCACCGAGCGCGGACCGTCTCGTCACT 14759
QY 1121 --GlnIleGlyAlaIleSerGln-----GlyMetSerValGlnL 1133
Db 14760 CCTCCGCGCGCGCGTACCGGAGGACCGTCTTCGCGAGATCGGCTTCGACTCGCTGA 14819
QY 1133 euHisValProTyrSerGluHisAlaLysAlaProValGly-----ProValThrM 1150
Db 14820 CGCGCTGCAACTGCGCAACGACTCGCGCGCGCCCGCTCGCTCGCGAGACCA 14879
QY 1150 etGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGln 1170
Db 14880 T-GGTGTTCGACTACCC--GACCCGCTGAGCTGGCCAGTATCTCGCGCGCGAGATG 14935
QY 1170 luGlnLeuSerProArgGlyGlnAlaGlyPro 1180
Db 14936 GTCGGCTCCGTCTGGAGGTGCG--GGCGCCCGTCCGCCACCGCGGTACCGACGAGAA 14992
QY 1180 ----- 1180
Db 14993 CCNATCCCATATCGCATGAGTGCCTACCCCGCGGAGTCACTCCCGGAGCAG 15052
QY 1181 -----ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyT 1198
Db 15053 CTGTGGGACCTGGGCTCTCG-GCACCGAGCGGATCACCGACTTCC-----CGGTCAA 15105
QY 1198 hrAlaLeuGlySerValProGly 1205

Db 15106 CGCGGCTGGAAACACCGCGCGGCTCTACGACCCCGAGCCCGGACCACTCTGGGACCACTTA 15165
QY 1206 -----GlySerIleThrLysGlyIleProSerThrArgValProSerAspS 1221
Db 15166 CTCACGCGAGGCGGCTCTCTGACGAGCGGAGGAGTTGACCCCATGTTCTTCGGCAT 15225
QY 1221 erAlaIleThrTyrArgGlySerIleThrHisGly----- 1232
Db 15226 CTGCCCCCGAGGCGCTGTATGACCCCGCAGCAGCGGCTCTCTCTGGAGACCACTG 15285
QY 1233 -----ThrProAlaAspValLeuTyrLysGlyIleThrA 1245
Db 15286 GAGGCGCTTCGAACGCGCGGACTCACCC-CACACCTC-----CGCGAGCGCTGACCG 15341
QY 1245 rgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProL 1265
Db 15342 GCACCTTCATCGGCTCCAGCTACAGGAGTACGGATGGCGCGCGACGGC--GCCG 15398
QY 1265 ysGlyHisVal-----IleTyrGluGlyLysLysGlyHisValL 1278
Db 15399 AGGCGCACCTGTGTACCGGACCGAGCGGTCTCTCCGCGCGCTCGCTATGTCT 15458
QY 1278 euSerTyrGluGly---GlyMetSerVal---ThrGlnCysSerLysGluAspGlyArgS 1296
Db 15459 TCGGCTGGAGGCGCGCGGTCAACCGCTGACACCGCTCTCTCTCTCTCTCTCTCTCTCTCT 15517
QY 1296 exSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluG 1316
Db 15518 CTGCACCTGGCTGCCA-----GGCGCTGGCAACGGGAGAGCACTCGCGGTGCG 15571
QY 1316 lyArgValGlyArgAlaIleSerSerAlaSerIleGlyLeuMetGlyArgAlaIleP 1336
Db 15572 GCGCGCGCACCGCTCATATGACCAACCCCAACGCGTTCGTGCGGTTCAGCGCGACGCGCC 15631
QY 1336 ro-----ProGluArgHisSerProHis 1344
Db 15632 CTGCGCGAGACGCGCGGTGCAAGCGTCTTCGAGAGCGCGCGCATGACCTCGGCC 15691
QY 1344 isLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrV 1364
Db 15692 GA---AGGTGTGGCATCTGCTCGTGGAGCGCTGTCGAGCGCGCGCGCAACGGCCAC 15748
QY 1364 alGluAlaGlnGluAsp---TyrLeuArgArgGlu---AlaLysLeuLeuLysArgGluG 1382
Db 15749 CCGGTCTCGCGGTGATCGCGCTCCGCCATCAACGAGACGCGCTCCCAACGCGCTC 15808
QY 1382 lyThrPro-----ProProProPro----- 1388
Db 15809 AGCGCCCCCAACGCGCCCCCTCCGAGAGCGGTCTCCGCGAGCGCTTCGCGAGCGCGG 15868
QY 1389 -----ProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysL 1407
Db 15869 GTGCGCCCCCGGAGATCGACCTCTCTGGA-----GGCCCAACGGCACC 15910
QY 1407 euLysProAla-----HisGluGlyLeuValAlaThrValLysGluAlaGly---- 1422
Db 15911 GGCACCGCGCTCGCGACCCCATCGAGC-----GCAGGCGCTG 15949
QY 1423 --ArgSerIle-----HisGluIleProArgGlu-----GluL 1433
Db 15950 TTCGCCACCTTACCGCGCGACCCCGTACCCCGAGACCGCGCTGTCTCTCTCTCTCTCTCT 16009
QY 1433 euArgHisThrProGluLeuProLeu----- 1441
Db 16010 TCGAACATCG--GCACTCCCATGTCGCGCGGTGTCGCGAGCATCATCAAGATGGTC 16066
QY 1442 -----AlaProArgProLeuLysGluGlySerIleThrGlnG 1454
Db 16067 ATGGCGCTGGCCACCGGTGTGATGCCCGCAGACCTTCACGCGCAGCGCTC----- 16119
QY 1454 lyThrProLeuLysTyrAspThrGlyAlaSerThrGlySerLysLysHisAspVal- 1473
Db 16120 --CAGCCACGTCTGACTGGAGCCCGGACCGGTACGCTGCTCGCGAGAACACCGACTGG 16177


```

/*tag= 1
/product= "Cytochrome P450 encoded by S. nodosus amphiN
gene"
complement(63250..64308)
/*tag= m
/product= "NRP-sugar aminotransferase encoded by S.
nodosus amphiDII gene"
complement(64324..65775)
/*tag= n
/product= "Glycosyl transferase encoded by S. nodosus
amphiDI gene"
/transal_except= (pos:65773..65775, aa:Met)
66081..70319
/*tag= o
/product= "Polyketide synthase multienzyme housing
loading module encoded by S. nodosus amphiA gene"
70366..79938
/*tag= p
/product= "Polyketide synthase multienzyme housing
extension modules 1 and 2 encoded by S. nodosus by amphiB
79956..112709
/*tag= q
/product= "Polyketide synthase multienzyme housing
extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
nodosus by amphiC gene"
70200297082-A2.
W0200297082-A2.
05-DEC-2002.
27-MAY-2002; 2002W0-1E000071.
31-MAY-2001; 2001IE-00000527.
(YUDU-) UNIV COLLEGE DUBLIN.
Caffrey JP;
WPI; 2003-201271/19.
P-PSDB: AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
AAE36129, AAE36130, AAE36131, AAE36132.
XX
Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
for preparing amphoterin derivative or analog antibiotic agent with
altered properties, in biosynthesis of polyketide other than
amphoterin.
Claim 1; Page 52-114; 276pp; English.
XX
The invention relates to the gene cluster encoding the polypeptides
responsible for the biosynthesis of the polyene antibiotic amphoterin
(amph) of Streptomyces nodosus. Polynucleotides of the invention are
useful for preparing amphoterin derivatives or analogue antibiotic
agents with altered properties and in the biosynthesis of polyketides
other than amphoterin. amphiDIII, amphiDII or amphiDI mutants are useful
for producing amphoterin derivatives glycosylated with alternative
sugars; amphiDII or amphiDI gene sequences are useful in engineered
biosynthesis of perosaminyl-amphoterinamide B; amphiDIII or amphiDII and
amphiN gene sequences are useful in the engineered biosynthesis of
perosaminyl-16-descarboxyl-16-methyl amphoterinamide B; amphiDIII, amphiDII
and amphiDI gene sequences are useful for preparing polypeptides capable
of addition of mycosamine to a polyketide other than amphoterinamide A or
B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX
The present sequence is S. nodosus amph biosynthetic gene cluster
SQ
Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 8.5e-11 Length: 113193
Score: 654.50 Matches: 644
Percent Similarity: 30.30% Conservative: 283

```

```

Best Local Similarity: 21.05% Mismatches: 1086
Query Match: 4.95% Indels: 1060
DB: 7 Gaps: 140
US-09-522-753-5 (1-2517) x AAD54645 (1-113193)
QY 39 LeuLeuGluTyrGlnHisHisSerArgAspTyr-----AlaSerHis--- 52
DB 11581 CTCCTTCGGCATCAGCGGCACCAATGTCACACCGTCTGGAGCAGCGCGGCACAC 11640
QY 53 -----LeuSerProGlySerIleIleGln 60
DB 11641 CGTCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11700
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyGlnArgSerGln 80
DB 11701 CCGACCCCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11742
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
DB 11743 -----CACCCACCTCCAGAACCCACCCGAGCGCGCT----- 11775
QY 101 GluPheIleGluSerLysArgProArgLeuLeuLeuLeuProAspProLeuLeuArgPro 120
DB 11776 -----CGCGACCTCG-----CCACTCCCTGGCCACCCACCGCG 11808
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 11809 CTCCTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11844
QY 141 LeuThrGlyLys---LeuGluProValSerProProSerProPro-----HisThrAsp 157
DB 11845 GG-ACCGGAAGGACTGCTCGCTCCCTCGCTCCCTCGCGCGCGCGCGCGCGCGCG 11903
QY 158 ProGluLeuLeuValProProArgLeu-----SerLys 169
DB 11904 CGCTCGTCCGAGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11963
QY 170 -GluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnI 189
DB 11964 GCAGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12023
QY 189 nileSerLys-----Le 193
DB 12024 CGTTCAGCAGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12083
QY 193 uLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysPro----- 207
DB 12084 TGTTCCCGCGCGAGGCGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12143
QY 208 -----ProGluProGluLysProValSerProProIle 219
DB 12144 CCGCGCTGTTCGCGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12203
QY 219 eGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaG 239
DB 12204 CCGACTTCGTTCACCGGTCACTC-----CATCGGTGAGATCGCGCG 12242
QY 239 uAlaAlaHisArgIleLeuGluGluGluGluGluGluGluGluGluGluGluGluGlu 259
DB 12243 CCGCGCACATCGC-----CGGGGTGTTCACCTCCAGGACGCGCGCGCGCGCG 12296
QY 259 nProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLy 279
DB 12297 CCG-----CGCACCTCATGAGGCGCGTCCGTC- 12327
QY 279 sLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCy 299
DB 12328 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12360
QY 299 sGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAs 319
DB 12361 -----GGAAG 12365

```

```
Db 37347 -----ATGCCGCGCTTGGCGCGTGGCGCGCTTGGCGCGTGGCGCTTGGCGCGC 37295
Qy 2111 -----GluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyVa 2126
Db 37294 GTTCCGCGGTGCGCGCGTGGCGCGCGCGCTTGGCGCGGTGGCGCGCGCT 37235
Qy 2126 lLysGlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAs 2146
Db 37234 GCCC----- 37231
Qy 2146 pTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPh 2166
Db 37230 -----CCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37187
Qy 2166 eProGlyAlaSerCysProValLeuLeuAspLeuLeuArgProProSerAspLeuTyrLeuPr 2186
Db 37186 GCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37139
Qy 2186 oProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgse 2206
Db 37138 GCCATTG-----CCACCTGTGCGCGCGTCCGCG----- 37111
Qy 2206 rProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValse 2226
Db 37110 -CCGTCCGCG-----TGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37085
Qy 2226 rProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTy 2246
Db 37084 TCCACCC-----TGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37042
Qy 2246 rArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrse 2266
Db 37041 -----TGTCCACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36998
Qy 2266 rGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerly 2286
Db 36997 GTCCCGCGCGGTG----- 36985
Qy 2286 sLysGlnGluIleAsnLysLysLeuAenThrHisAsnArgAsnGluProGluTyrAsnIl 2306
Db 36984 -----CCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36971
Qy 2306 eSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetTh 2326
Db 36970 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36934
Qy 2326 rTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIl 2346
Db 36934 ----- 36934
Qy 2346 eArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluSerProProLeuSerAlaAs 2366
Db 36933 -----CCGGTGGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36893
Qy 2366 nAlaPheAsnProLeuAsnAlaSerLeuProAlaAlaMetProIleThrAlaAl 2386
Db 36892 GCCG-----CCGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36840
Qy 2386 a-----AspGlyArgSerAspHisThrLeuThrSerProGlyGlyGl 2400
Db 36839 CGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36780
Qy 2400 yGlyLysAla-----LysValSerGlyArgProSerArgLysAlaLy 2415
Db 36779 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36720
Qy 2415 sSerProAlaProGly-----LeuAlaSerGlyAspArgProProSerVa 2430
Db 36719 CTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36660
Qy 2430 lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
Db 36659 TGCCGCGCGCGTGTAGTACTGAGGAGGCTGCGC-CGGCGCGCT-CCGGTG----- 36616
```

```
Qy 2450 pGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetAr 2470
Db 36615 -----CCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTA----- 36574
Qy 2470 gLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyPr 2490
Db 36573 -----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 36539
Qy 2490 oLeuAlaGlyPro 2494
Db 36538 ACCGCGCACCGCG 36526
RESULT 42
AAD54645
ID AAD54645 standard; DNA; 113193 BP.
XX
AC AAD54645;
XX
DT 26-JUN-2003 (first entry)
XX
DE Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster.
XX
KW Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
XX
OS Streptomyces nodosus.
XX
Key Location/Qualifiers
CDS complement(4..1824)
FT /*tag= a
FT /product= "ABC transporter encoded by S. nodosus amphG
FT gene"
FT /*tag= b
FT /product= "ABC transporter encoded by S. nodosus amphH
FT gene"
FT /*tag= c
FT /product= "GDP-mannose dehydratase encoded by S. nodosus
FT amphDIII gene"
FT /*tag= d
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
FT nodosus amphI gene"
FT /*tag= e
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 15, 16 and 17 encoded by S. nodosus
FT amphJ gene"
FT /*tag= f
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 18 and thioesterase encoded by S.
FT nodosus amphK gene"
FT /*tag= g
FT /product= "Cytochrome P450 encoded by S. nodosus amphL
FT gene"
FT /*tag= h
FT /product= "ORF1, hypothetical protein"
FT complement(58756..59610)
FT /*tag= i
FT /product= "ORF2, hypothetical protein"
FT 59869..61470
FT /*tag= j
FT /product= "ORF3, hypothetical protein"
FT /transl_except= (pos:59869..59871, aa:Met)
FT /*tag= k
FT /product= "Ferredoxin encoded by S. nodosus amphM gene"
FT complement(62051..63250)
```

```
QY 1476 -----LeulleGlySerProGlyArgThrPhe----- 1484
Db 38927 AGGGGATCCCGACGTGTTTCAGCGCTCTGCTGTCGGGGTGCAGATCCGCGTTACGGCC 38868
QY 1485 -----ProValHisProLeuA 1491
Db 38867 AGCAGTGGCCCGCGTGTGCAGCCCGCGGGTGTGTCGGCGCTGTTGGCCGAGTTGG 38808
QY 1491 spValMetAlaAspAlaArgAlaLeuGluAArgAlaCysTyr----- 1504
Db 38807 ACTCGTTCGTCGCTCC-----GAGATCGCGGGCTACAAAGTCGCGCGCAGTCTGT 38757
QY 1505 -----GluG 1506
Db 38756 GGTTCGTCAGCAGGTGAAGCTTCGCCCGCGGTAAAGCGGACTACCGTCGGGCCAAGG 38697
QY 1506 luSerLeuLysSerArgProGlyThrAlaSerSerGly-----GlySerI 1522
Db 38696 AGCAGACCGAGCGCGGCCAGCGATGACGTGATCGCGGACACGTGACGAGCGGTTCCT 38637
QY 1522 leAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuT 1542
Db 38636 GATTGCGCGGTCCGCG- GCTATGCGCGCGGTGCGGGATACCGCTGGGTCCCGCTTG 38578
QY 1542 hrTyrGluAspHisGly-AlaProPheAlaGly----- 1552
Db 38577 TCGTCAAAATCT--GGCTGGACCACTCGCGCGGACCGCGCGCACCATCGCGCGCGC 38521
QY 1553 -----HisLeuProArgGlySerProValThr 1561
Db 38520 TGGGCCCATCGCACTAGACCGCGCGCTGCGCCCTTGC CGCGCTCGCGCGTCCGCG 38461
QY 1562 MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGln 1581
Db 38460 CCGCGCGCGCGTGCCA-----CCGCTGGTGCCTTACCACCGCTCCGCGCGCACCG 38407
QY 1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db 38406 CCGCTGCGGTGTCCTCCCTTGC CG-----CCATTCCACCGCGTGCCG 38365
QY 1602 GluHisProHisProLysProTyrGluHisLeuLeuArgGlyValSerGlyVal 1621
Db 38364 CCGTTCACCAACACCG----- 38347
QY 1622 AspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIle 1641
Db 38346 -----CCGGTCCGCCATCGCGCGCGTGC CGCGCGCGAGCGGTG 38308
QY 1642 ProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyr 1661
Db 38307 CCACTGCG-----TTGCCGCGGTGCACCGCTTG 38278
QY 1662 ProHisLeuTyrPro-----ProTyrLeuIleArgGlyTyrProAspThrAla 1677
Db 38277 CCGCGGTGTGCGCGCGCGCGCAGCACCGCGTCCCGGGGCGACTCCGGGCTCGCG 38218
QY 1678 AlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHis 1697
Db 38218 ----- 38218
QY 1698 AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu 1717
Db 38217 -----TCGCG----- 38212
QY 1718 SerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnVal 1737
Db 38211 -----CCAATGCGCGCGGTG 38197
QY 1738 ProHisLeuProValLeu-----ValProProThrProGlyThrProAlaThrAlaMet 1755
Db 38196 CCGCGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38137
QY 1756 AspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSer 1775
38136 CCGCGCTTCGCCGCAATTCGCCGCGCGCACCG----- 38107
QY 1776 ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db 38106 CCGCTGCCCGCG-----GTCCCGCGGTGCC- GCGGATCGCGCGAC 38066
QY 1796 ArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLeu 1815
Db 38066 ----- 38066
QY 1816 ThrSerThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSer 1835
Db 38065 -----CCGGGGTTCGCCGCGCGC 38045
QY 1836 GlySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaSerHisSer 1855
Db 38044 GTTCGCCCGGTGCCCGCGGTGCCCGCG-----GCCCGCTCCCGCGCGGTGCC 37991
QY 1856 HisAlaHisGlnHisSerProLysSerProArgThrGlnAspAlaLeuGlnArgPro 1875
Db 37990 GCGGTGCCCGCATTCGCCGCGCACCGCGGTGCCCGCGCGCGCGCGAC 37931
QY 1876 SerValLeuHisAsnThrGlyMetLysGlyIleThrAlaVal-GluProSerLysPr 1895
Db 37930 CTGCCCGCTTCGCCCGGTACC-----GCCGTTGCCCGCTGCCCGCATCGCGCC 37877
QY 1895 o---ThrValLeuArgSerThrSerSerProValArgProAlaAlaThrPhePr 1914
Db 37876 AGTTCGCCCGCGCAGCTGCCGATGCCCGCGCGGTGCCCGCGGTGCCCGGTGCC 37817
QY 1914 oProAlaThrHisCysProLeuGlyThrLeuAspGlyValTyrProThrLeuMetG 1934
Db 37816 GCGGTGCCCGCTTCGCCCGCGGTGCTCAGTGTGAGTGCGCGCGCGCGCGCGCGCG 37760
QY 1934 uProValLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAs 1954
Db 37759 CCG-----TTGCCCGCATTCGCCCGCGCGCGCTTCGCCCGCGGTGCCCGCGCG 37706
QY 1954 pThrGlyHisAlaPheLeuAlaLysProAlaArgSerGlyLeuGluProAlaSerSe 1974
Db 37705 AGCG-----GTCAAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37661
QY 1974 rProSerLysGly----- 1979
Db 37660 ACCGTCCCGCGCTTGCACCGCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37601
QY 1979 rGluProArgProLeuValProValSerGlyHisAlaThrIleAlaArgThrProAl 1999
Db 37600 TGTCTCGTTGCCCGCTTACCGCGCTCACCGCGCGATGCCACCGCTGCCCGCGGTGCC 37541
QY 1999 aLysAsnLeuAlaProHisHisAlaSerProAspProAlaProAlaSerAlaSe 2019
Db 37540 AGTACCGGTGAGCGCG-----CCGTCCCGCGCGGTGCCCGCGGTGCCCGCGCG 37493
QY 2019 rAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnLeuLeuLeuAr 2039
Db 37492 GCGCGCG----- 37486
QY 2039 gSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProVa 2059
Db 37485 -----TTCGCGCG-----GCACCGCGCGTCCCGCT 37460
QY 2059 lSerSerProSerLeuThrHisAspLysGlyLeuPro-----LysHisLeuGluLe 2077
Db 37459 GCGGTGCCGTGTTGCCCGCGCGCGCGGTGCCCGCGGTGCCCGCGGTGCCCGCGGT 37400
QY 2077 uAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuG 2097
Db 37399 GCGCGGTGCCGTGCCGTGCCGTGCCGTGCCGTGCCGTGCCGTGCCGTGCCGTGCCGT 37348
QY 2097 yGlyGluAlaAlaHisLeuProHisLeuArgProLeuPro----- 2110
```



```
QY 152 SerProHisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGlu 171
Db 42299 TCACCCCGCATG-----CCCGGTTGCTCGCGGTGCGCGCGGTG----- 42258
QY 172 LeuIleGluAsnMetAspArgValAspArgGluIleThrMetValGluGlnIleSer 191
Db 42258 ----- 42258
QY 192 LysLeuLysLysGlnGlnLeuGluGluAlaAlaLysProGluProGlu 211
Db 42257 -----GCTAGCGCGCGCGCGCG 42237
QY 212 LysProValSerPro-----ProProIleGluSerLysHisArg----- 224
Db 42236 TTGCCCGCGCGCGCGCTGTGCGCGCGCTGCGCGCGAGTCGTCGCGCGCGCGCG 42177
QY 225 -----SerLeuValGlnIleIleTyr 231
Db 42176 GTCCCGCCTTTGCCCGCGGTGCGCGGTGCGCGCGGTGCGCGGTGCGCGGTG----- 42123
QY 232 AspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGln 251
Db 42122 -----CTGAGTCGCTG 42111
QY 252 ValGluLeuPro-Leu----- 256
Db 42110 GTGTTTAGCGCGCTTTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGCG 42051
QY 257 -----TyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIle 270
Db 42050 TTGAAGTTGGCCCCCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 41991
QY 270 eLysIleAsnGlnAlaMet---ArgLysLysLeuIleLeuTyrPheLysArgArgAsnHi 289
Db 41990 CCGGTGCGCGCTTTGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 41931
QY 289 sAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLe 309
Db 41930 GTCCCGCCACACCGCGCGTGC CGCGGTGCGCGCGGTGCGCGGTGCGCGGTGCGCG 41871
QY 309 uGluLysLeuValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysVa 329
Db 41870 CAGACACCCCGCGCGCGCTTGTCCCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 41811
QY 329 lArgGlu-----TyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLe 346
Db 41810 CCGCGCGCACCGTGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 41757
QY 346 uGlnGluArgMetGlnSerArgValGlnArgGlySerGlyLeuSerMetSerAlaAl 366
Db 41756 CCACCGCGCAACCGGTGCGCGGTGCGCGGTGCGCGCGGTGCGCGGTGCGCGGTG 41718
QY 366 aArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGl 386
Db 41717 TCGCGTGGCGCGCGCGCGCGCGCGCTC-----CGCGGTACCGCGCTTGG 41673
QY 386 uLysProGlyAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCy 466
Db 41597 CCGCGCGCGCGCTCGCGCGCTCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCG 41538
QY 466 sValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSe 486
```

```
Db 41537 CCGGTGGGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 41478
QY 486 rTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db 41477 C---CGCGCGCGCGCGCGCGGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 41421
QY 506 nGlnGlnGlnProMetProArgSerSerGlnGlnGluLysAspGluLysGluLysGl 526
Db 41420 CCGCGCGCGCGCGCGCG-----CGCGGTGCGCGCGCGCGCGCGCGCGCGGTG 41370
QY 526 uLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeuLe 546
Db 41369 GCGCGCGGTGTCGCGCGTGC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 41336
QY 546 uLysGluLysThrAspAspThrSerGlyGluAspAspGluLysGluAlaValAlaSe 566
Db 41336 ----- 41336
QY 566 rLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMe 586
Db 41335 -----CGGCTCCGCGCGTACCGCGCTTGGCGCGCGCGCGCGCGCGCGCG 41283
QY 586 tAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSe 606
Db 41282 G----- 41282
QY 606 rMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGl 626
Db 41281 -----TGGTCCGATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 41247
QY 626 yLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVa 646
Db 41246 GCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 41196
QY 646 lSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluLeLe 666
Db 41195 CCGCGCGTCCCGCGCGATGCGGTGGGTGTTATCGCGCGCGCGCGCGCGCGCG 41153
QY 666 uGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLys-A 686
Db 41152 -CGCAGCACCC---CGCGCGCGCTTGTTCGCGCGGTGCGCGCGCGCGCGCGCG 41097
QY 686 lAProAlaAlaLaserGluGluAlaAlaPheProValValGluAspGluMetG 706
Db 41096 TTGGTCCCGCGCACCGCTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 41053
QY 706 lAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisA 726
Db 41052 --TCCCGCGGTGCGCGCGC-----GCCGCGGAGCGCGGTGGTG 41016
QY 726 lAsrGlyAsnGluValProArgGlyGluCys-SerGlyProAlaThrValAsnAsnSer 745
Db 41015 CCGCGCGCACCGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40959
QY 746 SerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn 765
Db 40958 -----CGGTGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 40923
QY 766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro 785
Db 40922 GCGCGCGCGCGCGCGCGCT-----CGCGGTACCGCGCTTGGCGCG 40884
QY 786 ProArgThrSerArgAla-----ProIleGluProThr 797
Db 40883 CCGCGCGCGCGCGCGCGGTGCTGTGGTGGCGATCATGCGCGCGGTGCGCTCCGCGTG 40824
QY 798 ProAlaSer---GluAlaThrGlyAlaProThrProProProAlaProProSerPro--- 815
Db 40823 CCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40764
QY 816 -----SerAlaProProValProValProLysGluGluLys 827
```


1775 SerProLeuSerProGly----- 1780
1780 AGACCTCGCGCGCGGAGCCGATGTCGAGAGACCGCTGGAGGAGCGGTATCGAGCTCC 7003
1781 -----GlyProThr 1783
1782 TGACCGCGTCCGTGGAGGACACAGACCGAGAGAGAGAGAGCGGTGGTGTCTCTCTCC 6943
1784 HisLeuThr----- 1786
1786 CGCTGACTCCATGGAGGCTTGAGCAGTTCCGCGACCTCCGCGACCGCGCGCGCATGG 6883
1787 -----LysProThrThrThrSerSerSerGluArgGlu 1797
1797 GCACCTCAAGCGCGGAGCGAGGACCGCTCGGACACACCGTTGTCAGTCCGTCGA 6823
1798 ArgAspArg----- 1806
1806 GGAGCTCGCGTGGTGGTGGGCGGTCGACGAGGAGGAGCGACCGATGGAGTGAGGCG 6763
1807 AspArgGluArgGluSerHisLeuThrSer---ThrThrThrValGluHisAlaPro 1825
1825 GAGCGGGGAGGATTCGGTCAGGCTTCCAGGTGACCGGTAGCGCGGATCGG--- 6706
1826 IleTrp----- 1827
1827 TCGGGCTCTGGTGGTCTCGCGGGGTGGTGGGGGTGGTGTCTGAGCCAGTAGTGGT 6649
1827 ----- 1827
1828 GGTGGTTGAAGGGTAGTGGGAGGGTGGTGGTGGCTGGGGGAGGGCGGGTCCC 6589
1828 -----ArgProGlyThrGluGlnSerSerGlySerSerGly 1839
1839 AGTGAGGTGGCGCGTGGCGCCAGCGCTCGCGGTGGAGGTACAGAGCGGTGGCGGC 6529
1840 SerSerGlyGlyGlyGlySer-----SerSerArgPro 1851
1851 TGTGCTCTGGCGCGCAGGCTCCGGTGGCATGGACACTCCGGTCCGCGGCTCGCGCG 6469
1852 AlaSerHisSerHisAla----- 1857
1857 TCTGCTCGATCGATGACGAGTACCGGTGGGCGACTGACCTCGATGAACCGAGTGGC 6409
1858 -----HisGln-His-----Se 1861
1861 CCTCTCCAAACAAACACACACCGCGGAGAACTCAACCGGTCCGCGAGATTACGGA 6349
1861 rProIle----- 1863
1863 ACCAATACCGCATCGAAACCGGCTCGCGCGGCTGCTCTCGCGGAGCGGTGGAACACA 6289
1864 -----SerProArgThr-----GlnAspAlaGlnGlnArgProSe 1876
1876 CCGGACCGCGGAGTACCGGGGCGGATACCGAGACAGACCGCTCCAACTCGCCCTTCA 6229
1876 rValLeuHisAsnThrGlyMetGlyGlyIleThrAlaValGluProSer----- 1893
1893 GGTCTCTCCATGCGCGGAGTGGAGCGGTAGTCCACCGGTATCAACCGAGCTGCACAC 6169
1894 -----LysProThrValLeuArgSerThr-- 1901
1901 CCTCGGCTGCAGACGCGGCACATACCGGCGCACCGCGCGCGATCCC-CGGAACACAC 6110
1902 -----SerThrSerSerProValArgPr 1909
1909 GTGAGCGCGGAGCTTACCGCGGACCTCCACACCGCGGCGCATCTCGCCAGCCGCGCC 6050
1909 o-----AlaAlaThrPheProAlaThrHisCysProLeuGlyGlyTh 1924
1924 TCGACCTCGCGGCGAGCGGAGACAGCGCGCATCGCACACGAGCCG---CCAGATAC 5993
1924 rLeuAspGlyValTyrPro-----ThrLeuMe 1933

5992 CGGCCAATCAACCCCGCGCGAGAGCGACACAGCGCGCGCATCTCTCCAGCGCAACGCC 5933
1933 tGluProValLeuLeuProGlyGluAlaProArgValAlaArgProGluArgProArg-- 1952
5932 CCGCGCACCGTGGCAGCAGGATCTCCCTCGGANTGATCCACACCGAGCGACGTCGCC 5873
1953 -----AlaAspThrGlyHisAlaPheLeuAlaLysProAlaArgSerG1 1968
5872 ACACCATCGCTGCGAGTACCGGCGCAGCACCATCACCGCCCGCTCAGGGCT-- 5815
1968 yLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProVa 1988
5814 -----GCACCATATCCACCGCTCCACCGCGCGCGCGCGCGACCATCTCCAGC 5765
1988 lSer-----GlyHisAlaThrIleAlaArgThrProAlaLysAsn-- 2001
5764 AAATCCCACTCCACATGAACACTCAGCGCGCGCGCACACTCCCGCATCGACGCGCAAC 5705
2001 ----- 2001
5704 ACCTCGAAGACCCAGCAGCGCGCGCACCATCCCAACCATCTCGCTCCCTGACGAGG 5645
2001 ----- 2001
5644 AAACCAACACCAACCGCGCGCGCACCGGACCGCGTCTCGGGTCCACCAACCGCGCGGTG 5585
2001 ----- 2001
5584 ACCACACCGCGTCCGGATCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5525
2001 ----- 2001
5524 TCACCG 5465
2002 -----LeuAlaProHisAlaSerProAspProPro-- 2012
5464 GCACATCCCG 5405
2013 -----AlaProProAlaSerAspPro-----Hi 2022
5404 TCCCG 5345
2022 sArg-----GluLysThr-----GlnSerLysProPheSerIleGlnGluLeuG1 2037
5344 TCAGGACCG 5285
2037 uLeuArgSerLeuGlyTyrHisGlySerSer----- 2047
5284 TCGCGCGGTTCGGGGGCTTCGAGGATCAGATGGCGGTGGTACCTGACACCAAGGCG 5225
2048 -----TyrSerProGluGlyValGluProValSerProValSerSerProSerLe 2064
5224 GAGACCG 5165
2064 uThrHisAspLysGlyLeuProLysHisLeu-GluGluLeuAspLysSerHisLeuGluG 2084
5164 CACACACCG 5112
2084 lGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuP 2104
5111 -----CGTCCGCGCGCAACCA-----CCCATACCGCATCGCGCAACACCATCTTG 5069
2104 roHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaP 2124
5068 ATCAGAC-CAGCCACCG 5010
2124 roGlyValLysGlyHisGlnArgValThrLeuAlaGlnHisSerGluVallet 2144
5009 CA-----ACCA 5004
2144 hrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuT 2164

8927 GGCGTACTCTCTGAGGTGAGCCGAC-----GTATAC 8895
1226 ArgGlySerIleThrHisGlyThrProAlaaspValLeuTyrLysGlyThrIleThrArg 1245
8894 CCCGGTGC-----GTCGGTGCAGCA 8871
1246 IleIleGlyGluaspSerProSerArgLeuaspArgGlyArgGluaspSerLeuProLys 1265
8870 GTG-----CGGTGCATACCGCCGCTCGATCGCTCCAGCA 8832
1266 Gly-----HisVal-IleTyrGluGlyLysGlyHisValle 1278
8831 GGTCTCAGCAACAGCCGCTGCTCGGATCATCGCCAGCCTCAGCGCGAAATCC 8772
1278 uSerTyrGluGlyMetSerValThrGlnCysSerLysGluaspGlyArgSer-SerS 1298
8771 GAAGAACTCCGCGTCGAACCTCCGCGCTCGTGAGAAACCCCTCCCGCATATGT 8712
1298 erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV 1318
8711 CTTCCCGCCACACCCCGCTCCGGTCTGAGCGGAATCCAGGTCCAGCCCGCGTGGT 8652
1318 alGlyArgAlaIle-----SerSerAlaSerIle----- 1327
8651 GGGCGCGCGCTATGACGTACGCTCTCGTCCAGCATCTCCACAGCGCTCAGGGT 8592
1328 -----GluGlyLeuMetGlyArgAlaIleProProGluArgHisSerP 1342
8591 GCCCGCCCCACCGGGAAACGGCCATGCCAGCATGACCACCGGGTCTGTCGCG 8532
1342 roHisHisLeuLysGluGlnHisIleArgGlySerIleThrGlnGlyIleProArgS 1362
8531 CACGG----- 8527
1362 erTyrValGluAlaGlnGluaspTyrLeuArgArgGluAlaLysLeuLysArgGluG 1382
8526 -----CGACGGCGAGTGGCGCTCATCGCGCG-G 8497
1382 lyThrProProProProSerArgaspLeuThrGluAlaTyrLysThrGlnAlaL 1402
8496 GCGCGGGGTCCCGCGCGCTGAGCGGTGAGCAGCATGTGGCCAGGGCGGTGCGGC 8437
1402 euGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaG 1422
8436 TGGGTGTCGAGATG-----ACGTGCGCGGTAGCT 8404
1422 lyArgSerIleHisGluIleProArg-----GluGluLeuArgHisThrProG 1438
8403 TCAGGCTGTGCGGCATTGAGCGGTGCGGAGTTCAGGAGGTTCAGGAGTCGAAAC 8344
1438 luLeuProLeuAlaProArgProLeuLysGluGlySerIle-----ThrG 1453
8343 CGGCGCACTTGAACGGCGGTTCGCGCGCATGCGTCCGTTCGGATGCGCGAGCAAA 8284
1453 lnglyThrProLysTyrAspThrGlyAlaSerThrTyrGlySerLysHisAspV 1473
8283 AGCGCGCTCGGTACG-----ACGAGTTCAGCAGCTTCGCGAGCGGTTCGG 8236
1473 al-----ArgSerLeuIleGlySerProGlyArgThrPheProp 1486
8235 TCTGTTTCAGCATGGAACACCGCGCGCAGCTCGGCGCGCGCTCCGCGGCTC 8183
1486 roValHisProLeuaspValMetAlaaspAlaArgAlaLeuGluArgAlaCysTyrGluG 1506
8182 -----GCGGAGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 8134
1506 luSerLeuLysSerArgProGly----- 1513
8133 TGTG-----ACACGCGCGCGGAGCGGATGCGGTGCGGTGCGGTGCGGTGCGGT 8077
1514 -----ThrAlaS 1516
8076 CGAGCTCGACAGTACGACGACGACGTGTTCTGCTGCCATGCGCGCCCGAGGCGGTGCA 8017

1516 erSerSerGlyGlySerIleAlaArgGly-----Alap 1527
8016 CGCGCGGTGCGGTTCATGCGCGCGACGCTCGCGCGACACTCTCTCTCCCGCGAC 7957
1527 roValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisG 1547
7956 CCGCGCGCATGCT-----CCCGCGCCCAAGGCGCCAGGCGACGAGGTGCGCG 7906
1547 lyAlaProPheAla----- 1551
7905 GCAGTCCGCGCGCACGCGACGCTCGCGAGGCGCTCGAGCGCGGTTCGCGCGCGCT 7846
1552 -----GlyHisLeuProArgGlySerProValThrMetArgLup 1565
7845 ACACACCTCGCGCGGTTCCTCCCATGTGCGGTGAGCGGAGGAGAGACGACGCGCT 7786
1565 roThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspArgLysL 1585
7785 CCAACCTTTTATGTCGCGAGTCAGTTGTCAGCAGCTCCGCGCGCGAC----- 7736
1585 euThrSerThrProArgGluIleAlaLys-SerPro----- 1596
7735 --ACCTTCGCGCGCGACGCTCTCGAAGCTCTCCGCTGACGCTGTCATCACCGCT 7678
1596 ----- 1596
7677 GTCCAGAAATCCCGCGGTGTGAAGACGCGCTTCGCGGATACGCGGTGACGAGCGCG 7618
1597 -----HiserThrValProGluHisHisProHisProHisProHisProHisLeu 1614
7617 CGAGCGCATCACGCTCGCGCACGTCACAGCGCGCT---GCACGACCTCGCAGCCTATGTC 7561
1615 Leu----- 1615
7560 CTCGCACTTCCTCGCGAGTTCGCGCGCGCGCGCTCGGTTCGCGCGCGGTGCTGTA 7501
1616 -----ArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAla 1631
7500 GGACGAGCGTTCGCT-CCCTCGCGCGCGCGCTCGCGCGAGCGCGCTCGCATGCA 7442
1632 PheAspPro-----ThrSerIleProArgGlyLeuProLeuAspAlaAlaAla 1648
7441 CCGATGCGCGGTGATGACGACCGCTCGCGCGCGCTGCGCGT---GAGGCGCGCGCT 7395
1649 TyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProTyr 1668
7384 GCGCGCGC-CCGCGACAC-----CCGCGCGCTAGATCCC----- 7350
1669 LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleAsnAsp 1688
7349 -----CGAGGC----- 7344
1689 TyrIleThrSerGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAsp 1708
7343 -----GCGCAGCGCGACCTGCTCTCGCGCGCGCGCGCGCGCGAC 7302
1709 -----MetLeuArgGlyLeuSerProArgGlnSerLeuAlaLeuAsnTyrAla 1725
7301 GCGCGCAGCAAGCTCAGGCGCGGTGCT-TCGAGTCTCGCGCGCGCTCTCCGGGTCT 7243
1726 AlaGly-----ProArgGlyIleIleAspLeuSer 1735
7242 GTGGGCGCTGAGGTCTCGCGCGTTCGCTGCTCGCGCGAGATCGATCA 7183
1736 GlnValProHisLeuPro-----ValLeuValProProThrProGlyThrProAlaThrAla 1754
7182 AGCGCGCCAGATATCGGAGTTCCAGGCGCGCGCTCGTCCGAGCGCGCGCGCGCG 7123
1755 MetAspArgLeuAlaTyrLeuProThrAlaProPheProPheSerSerArgHisSer 1774
7122 CCGCGCGCTCTCGCGGTCTCTCCCGGAAACCGCGCGCGCTCTCGCGCGCGCGCG 7063

QY 694 aAlaPheProValValGluAspGluMetGluAlaSerGlyValSerGlyAseGl 714
DB 10946 CGGCTCTCTCGGCGACGGTG--GAACACACCGGACCGCGGACTCGGGGGGGAT 10890
QY 714 uGluGluMetValGluGluAlaGluAlaLeuHisAlaSer----- 727
DB 10889 ACCGACAGCAGCCCGCTCCAACTCGCCCTTCAGTCTCTCCACATGCGGGAGTGAGAGGC 10830
QY 728 ---GlyAseGluValProArgGlyGluCysSerGly----- 738
DB 10829 GTAGTCCACGGTATCAACGGAGCC---TGACACCCCTCGGCTGACAGCGGCCACATA 10773
QY 739 -ProAlaThrValAsnAsnSerSerAspThr----- 748
DB 10772 ACCGGCCACCGCGCGCGATCCCGGAAACACCGTGGAGCGCGGACCGTTCAGCGCGC 10713
QY 749 -----GluSer-IleProSerProHisThrGluAlaAlaLys- 760
DB 10712 TACCTGTACTCGGGCCACTTCGCCAGCGCGGCTCGACCTCGCGGAGGACGGGAAC 10653
QY 761 -----AspThrGlyGlnAsn----- 765
DB 10652 AGCGGCATCGCACAGGACCGCGCGGATACCGGCCAATCAACCCCGCGCGAGCGGAC 10593
QY 766 -----GlyProLysProProAlaThrLeu----- 773
DB 10592 CACAGCGCGCGCATCTCCAGCGACAAACCGCCCGCACCGTGGCAGCAGCATCTCCCC 10533
QY 774 --GlyAlaAspGlyProPro-----ProGlyProP 783
DB 10532 CTGGGAATGACCCACACCGCAGCCACGTTCACACCCATCGCTGCCAGTACCGGGCCAG 10473
QY 783 roThrProProArgArgThrSerArgAla----- 792
DB 10472 CGACACCATCACCGCCACGTACGGCTGACCGCTGCACCATCCACCGCTCCACCCGCCCC 10413
QY 793 -----ProIleGluProThrPro----- 798
DB 10412 GCCCGACACCATCTCCAGCAAAATCCCACTCCACATGAACACTCAGCGCGCGCACACTC 10353
QY 798 ----- 798
DB 10352 CGGCATCGAGCGCGCAACACACTTCGAAGACCCCGACGCGCCGCCACCCATCCCCACCCA 10293
QY 799 --AlaSerGluAlaThrGlyAlaProThrPro-----ProAlaPro- 812
DB 10292 CTGGTCCCTGTACAGGAGGAACCAACACACCGCCCGCCACCGACCGGCCCTCTCCGG 10233
QY 813 -----ProSerProSerAlaProProVal----- 821
DB 10232 GTCCACACCGCAGCGTGACACACCGCGGTCCGGATCACCCCGCCCAACCGCGCAA 10173
QY 822 -----ValProLysGluGluLysGluGluThrAlaAla 833
DB 10172 ACGCCGGCAAGCTCAGCTTCACCCCGAGCGCACCAACCGCCGATACCCAAACCGCGA 10113
QY 833 laAlaProValGluGluGluGluGluGluGluGluGluGluGluGluAla 853
DB 10112 CCGGCGCGCA-----TCGTACCGGCCACATCCGCGAGCCGCCACACCGCGC- 10064
QY 853 alAspThrGlyLysAlaGluGluProValLysSerGluCysThr-----GluGluAlaG 871
DB 10063 --TCCACCATCTCGGCAACCGGCGCGAGCTCCCGCACCAACCGCGGAGCGCGCG 10006
QY 871 luGluGlyProAlaLysGlyLysAspAlaGluAlaGluAlaThrAlaGluGlyAlaL 891
DB 10005 AAACACACCGCGCACACACCGCAGCGGTGTGCTCTTCGGGACTGCCAGAGCGCT 9946
QY 891 eu-----LysAlaGluLysGluGlyGlySerGlyArgAlaThrAlaLysSers 909
DB 9945 TCGCACCTCTCGGCTCGGCGCGCTTCGAGGGCTTCGAGGGCTTCGGTACGCTCGGTGACGT 9886
QY 909 erGlyAlaProGlnAspSerSerAlaThrCysSerAlaAspGluVal 929

DB 9885 CGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGGCTTCGAGG 9826
QY 929 laGluGlyGlyAspLysAsnArg---LeuLeuSerProArgProSerLeuLeuThrProT 948
DB 9825 CTTGGGGGCTTCGAGGATCATGAGCGGTGTGTACCA-----CTGACACCGA 9778
QY 948 hrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysG 968
DB 9777 A-GGGGGAGACCGCGCGCGCGCGCGCTCCACCCCGCGCTCCCGCGCTCGGTC 9719
QY 968 lNArg-----AlaAlaAlaAla----- 973
DB 9718 AGCAACCGCACACACACCGACCGAGTCCCATGCGGCTCGGCTCATTCACATGCAAC 9659
QY 974 -----ProIleGlnValThrLysVal-----HisGluProPro- 985
DB 9658 GTCCGGCGCAACACCCCATACC-GCATCGCAACCATCTTGATCACACCGACCCACCC 9600
QY 985 ----- 985
DB 9599 CGCAGCGCTCGAGCATGACCAATATTGACTTCAACGACCCCAACACCGGACACTC 9540
QY 986 --ArgGluAspAlaAlaProThrLysProAlaProProAlaPro----- 999
DB 9539 AGCATCAGATCCCGCCCATAGCTCCAGCAACGCTCGGCTCGATCGATCAGCCAG 9480
QY 1000 -----ProProGlnAsnLeuGlnProGluSerAspAlaProG 1013
DB 9479 CGCGTCCCGCTCCGATGCGCTCCACCATCCATCCATCGGCGCCGCTCAACCGCG- 9424
QY 1013 lNlNProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysG 1033
DB 9423 -----CATTCGCCAAG---CCTGACGAATCACCGCTCTCGCGACGACCACT 9378
QY 1033 luAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProCysTrpThrSerGlyL 1053
DB 9377 CGGCGCGTCAAAACCATTCGACGCGCTCTGATTTCAGCGCGCGACCCACACA----- 9325
QY 1053 euProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPro 1073
DB 9324 --CCACCGCAACACCGGAT-----GCCCAACCGCGCGGCGATCCGA 9285
QY 1073 erAlaPheSerTyAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaAl 1093
DB 9284 CAACCGCTCCACGACGACCAACCCCGCGCACCTCGGCCCAAGCGGTCCCAT----- 9235
QY 1093 rgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAlaL 1113
DB 9234 --CCGCGCATCGCGAAGCGCTTGACCGCCCATCGCGCGCAACCCCGCTCGCGCGA 9177
QY 1113 yehiSerProValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnL 1133
DB 9176 AAACCTCCAGA-----ACAATCGGGTGGCGGCATCATCCGTCAC 9138
QY 1133 euHisValProTySerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuP 1153
DB 9137 ACCACCGCCACCGCATGTCACTCGCCACCCCGCAACCGCTCGG----- 9091
QY 1153 roLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuS 1173
DB 9090 -----CGCGTGTGTCAGCGGCGACGACGACGACGACGACGACGACGAGTGC 9048
QY 1173 erProArgGlyGlnAlaGlyPro-ProGluSerLeuGlyValProThrAlaGlnGluAla 1192
DB 9047 GATGGAGCGAAGAGACCTTCAGCCCGCGGCTGAGAGATCGGCGCGGAGAGAGCGCT 8988
QY 1193 SerValLeuArgGlyThrAlaLeuGlySerVal-----ProGly 1205
DB 8987 CGCGGACTTCCCGTTCGAGCAGATAGCTTCGTACTCTCTCGGGGCGCTTCGTCGCGCGG 8928
QY 1206 GlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyr 1225


```
Db 65367 CGCTCAACACCCAGCACCCTCCGCG-----AAGAGACC 65335
Qy 2293 sLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluI 2313
Db 65334 ACACACACCTCTCCGCGCGCAGCAGC-----GGCGCACAAACCCCGCCACGAC-- 65283
Qy 2313 ePheAsnMetProAlaIleThrGlyThrGly-----LeuMetThrTyrAr 2328
Db 65282 -----ACCACCGGCACCGGAGACCCGCGCGTCCACCTTCCCGTTCGG 65239
Qy 2328 gSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgly 2348
Db 65238 CGTCAACGGGATCTCATCCAGAACCCACCGCGCAGCCACCATGTATCAGCGAGACG 65179
Qy 2348 sAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPh 2368
Db 65178 CTGACGACCGCAGCATCACCGCA---GAGGACAGTCCACCGTGTCAACA----- 65130
Qy 2368 eAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAla----- 2386
Db 65129 -----CCGTGACATACGCAACCAA 65110
Qy 2387 -----AspGlyArgSerAspHisThrLeu----- 2394
Db 65109 CCGGTGATCACCGCGCGGTCTCCAGCACCAACACCGCACCGCACCATCATCAGGATG 65050
Qy 2395 -----ThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyAr 2408
Db 65049 CGCGGCACAGCACCCCTCCACCTCACCACCACTCCACGAGAAACCGGACCTTCACCTG 64990
Qy 2408 gProSerSerArgLysAlaLysSerProAlaProGlyLeu-AlaSerGlyAspArgProp 2428
Db 64989 ATCATCGCAGCAGCAACAACTCAAGCGTCCGCTCCACCGCACCGCAACCAATCCCC 64930
Qy 2428 roSerValSerSerValHis-----SerGluGlyAspCysAsnArgArgThrProLeuT 2446
Db 64929 CGTCCGATACATACGTCCACCCCGCGCACCGCAACCGGACCGCACCAACCGCTCCGCGT 64870
Qy 2446 hrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnP 2466
Db 64869 CAGACCCACAGCCCGCGAGTAGCCCGCG-----CCAGACCCACACCCGC 64825
Qy 2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuPro- 2485
Db 64824 CA-----CATACAGTCAACCCACCAACCGACCGGACCGCAC 64789
Qy 2486 -----AlaGlySerGlyProLeuAlaG 2493
Db 64788 GCGCAGCCCGTCTCCAGCACATACACACAGTGTTCGCGAGGGCGCGCCCATGG---G 64732
Qy 2493 lyProHisHisAlaTrp 2498
Db 64731 GCGCGCGCGTCCATGG 64715

RESULT 40
AAT80413/c
ID AAT80413 standard; DNA; 43280 BP.
XX AC AAT80413;
XX AC AAT80413;
DT 27-FEB-1998 (first entry)
XX XX
DE Tylactone synthase gene cluster.
XX XX
KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX OS
XX OS Streptomyces fradiae.
XX Key Location/Qualifiers
XX CDS 816..14243
FT /*tag= a
```

```
FT FT
FT FT /transl_except= (pos: 816..818, aa: Met)
FT FT /note= "ORF1 encodes protein shown in AAW22601"
FT FT 14351..19945
FT FT /*tag= b
FT FT /transl_except= (pos: 14351..14353, aa: Met)
FT FT /note= "ORF2 encodes protein shown in AAW22602"
FT FT 20010..31199
FT FT /*tag= c
FT FT /transl_except= (pos: 20010..20012, aa: Met)
FT FT /note= "ORF3 encodes protein shown in AAW22603"
FT FT 31232..36067
FT FT /*tag= d
FT FT /note= "ORF4 encodes protein shown in AAW22604"
FT FT 36249..41774
FT FT /*tag= e
FT FT /note= "ORF5 encodes protein shown in AAW22605"
XX XX
XX EP791655-A2.
XX XX
XX 27-AUG-1997.
XX XX
XX 19-FEB-1997; 97BP-00301056.
XX XX
XX 22-FEB-1996; 96US-0012078P.
XX XX
XX (ELIL ) LILLY & CO ELI.
XX XX
XX Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX XX
XX WPI; 1997-418046/39.
XX XX
XX P-PSDB; AAW22601, AAW22602, AAW22603, AAW22604, AAW22605.
XX XX
XX DNA encoding Streptomyces fradiae tylactone synthase domain - for
XX XX production of tylosin-related polyketide compounds.
XX XX
XX Claim 2; Page 8-66; 220pp; English.
XX XX
XX This sequence represents the tylactone synthase gene cluster of the
XX XX invention. This sequence is also referred to as the tylG gene, and was
XX XX isolated from Streptomyces fradiae. This sequence encodes multifunctional
XX XX proteins which direct the synthesis of the polyketide tylactone, isolated
XX XX from Streptomyces fradiae. Tylactone is the basic building block of the
XX XX antibiotic tylosin. The DNA sequence can be modified so as to alter the
XX XX type of carboxylic acids incorporated, the number of carboxylic acids
XX XX incorporated and/or the post-condensation reactions performed, thereby
XX XX resulting in novel tylosin-related polyketides
XX XX
XX SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,48e-11 Length: 43280
Score: 661.50 Matches: 696
Percent Similarity: 28.75% Conservative: 275
Best Local Similarity: 20.61% Mismatches: 1111
Query Match: 5.01% Indels: 1310
DB: 2 Gaps: 138

US-09-522-753-5 (1-2517) x AAT80413 (1-43280)
Qy 54 SerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGln 73
Db 12977 AGCCGCGCGCGCGCGTCTCAGCAGCAGCAGATGCGCGCGCGCGCAGATG 12918
Qy 74 ProGly-----AsnGluArgSerGlnGlu 81
Db 12917 CCGGCGCAGCAGCGCGCGCGCGTCCGCGCGCGCGTCCAGCAGCAGCGCGCGCG 12858
Qy 82 LeuHisLeuArg-ProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetG 101
Db 12857 GTCGAATGCGGACCGCGTCTCTCTCCGCGGTCCCGCGCGGAGGAGGACCGCGGTCG 12798
Qy 101 uPheIleGluSerLysArgProArgLeuGluLeuLeuProAsnPro----- 116
```

Qy 1813 SerIleLeuThrSer----- 1817
Dy TCACCAACCGCTCGATACCGCGCTCGCGCAGTTCGATGACTCCACGCGCAGATCCACC 67296
Qy 1818 -----ThrThrValGluHisAlaPro-----IleTrpArg----- 1828
Dy TGTGTCGAGCAGCGCAACCACTTCGCGCAGCGCAGTTCGGCGTATCCCGACCGCAGC 67236
Qy 1829 -----ProGlyThrGluGlnSerSerGlySerSerGlySerGlyGlyGly 1845
Dy ACCCGTCGCGCGCCACAGACCTTCGACACCGCGCGCAGCGGTGAGTCCACCGGA 67176
Qy 1846 GlySerSerArg-ProAlaSer-HisSerHisAlaHisGlnHisSerProIleSerP 1865
Dy CGACACAGATCGAACACAGCTGCCACATACAGTCCAC-----CCCA 67131
Qy 1865 roArgThrGlnAspAlaLeuGlnArg-ProSerValLeuHisAsnThrGlyMetLys 1884
Dy CCAGAGGCCAACGCGCTGTGGAACAGCAGCGCCCTCTTCGAGGCGCGACACCGGCAGA-- 67074
Qy 1885 GlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSer- 1902
Dy ATCTCTCCGACCAACCGCGCAGCCCGCCCTTCGACAGCGGTGAGTCCACCTGG 67020
Qy 1903 ---ThrSerProValArgProAlaAlaThrPheProAlaThrHisCys- 1919
Dy TTCACTTCCACCGCTACGTCGAGCGGTGAGGCGCGCGCTTCGGGTGTCGGCGG 66960
Qy 1919 ----- 1919
Dy TGTTTCGACAGCCCTTCAGGCGCTCGTCCAGTAGCGGGCAGCTCCCGTATCCGGGAC 66900
Qy 1920 -----ProLeuGly--- 1922
Dy TCCGCGCAGCAGCTGTCGCGCACGACCATGTACGCCGCGCGCGCTTCGGGGTG 66940
Qy 1922 ----- 1922
Dy TCCAGCGTATCGGTTGAACTCGACCGCTGCGCCAGCGGCGAGTTCGGGGTCTGGCGG 66780
Qy 1922 ----- 1922
Dy GCTCCACCGCAACGCGCTTCGACCGGGGACCGAGTCCGTCATCCCTCGTCACCGGTG 66720
Qy 1922 ----- 1922
Dy TCGTCGTTCCGTTGTAAGCGCGCGAGGTAGTTGAACCGAACTCGGGGTGGCATGC 66660
Qy 1922 ----- 1922
Dy CGGCGCAGAGCAGACCGTCTCGGGATCGGATAGCGCAGGATGCCATACCGACGCC 66600
Qy 1922 ----- 1922
Dy TGTCAGGCAGCAACGAGAGTCTCTTCACTCAGCAGCACCACCGCCACCGCA 66540
Qy 1923 -----GlyThrLeuAspGlyValTyr 1929
Dy CCACCCGACGCACTCGTCCAGAGCGGACTTCCACCGCGCACACCGACCGGA---TAC 66483
Qy 1930 ProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGlu 1949
Dy ACTGTGTGAACCAACGACAGTACGAGACATCAGACCCCGCCAGCGCTCTCATTA 66423
Qy 1950 ArgPro-----ArgAlaAspThrGly-His-----AlaPheLeuAlaLysPr 1963
Dy CGCCGTCGCCCTCAACATCCACCAACCGCGCCACCGCGCGCCACCGCGCGAC 66363
Qy 1963 oProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgPr 1983
Dy GCCAGCCGGAATG-----CGCGCAGCAACACCTCTCTGCACACCAACACCGCAACCC 66312

Qy 1983 oLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn----- 2001
Dy GACGGCAGCTTCGTCAGCAACGCGCAGCGGACACCGACACCGCGCGCGCGCGCGCGCG 66311
Qy 2002 -----LeuAlaProHisHisAl 2007
Dy ACCACCCCGCGGAAACATCCACCGCAGCACCACGACGCGCGCGCGCGCGCGCGCGCGCG 66251
Qy 2007 aserProAspPro-----ProAlaProPr 2015
Dy GCACCCCGAACCATTCGCGGACCAACTCCACTTCCCGCGCAGCTCGCGCGCGCGCGCGCG 66191
Qy 2015 oAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnI 2035
Dy TGCTCCGCAACG 66131
Qy 2035 uLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPr 2055
Dy -----CACGGCCACCG 66073
Qy 2055 oValSerProValSerSerProSerLeuThrHis-----AspLysGlyLeuPr 2071
Dy GCAGATCCCGCGCAGCACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66038
Qy 2071 oLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPr 2091
Dy AGCACCAACACCG 65978
Qy 2091 oGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProG 2111
Dy CGGCG 65921
Qy 2111 uSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr 2131
Dy CGCGCAGCGCGCAGCATCCACTGACATCCACACCGCGCGCGCGCGCGCGCGCGCGCGCG 65882
Qy 2131 gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisH 2151
Dy -----ACCGGCAATCCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 65827
Qy 2151 sProGlnGlnLeuSerAla-----ProLeuProAla----- 2161
Dy TCCGGAACAACCATCAACACACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 65804
Qy 2162 -----ProLeuTyrSerPheProG 2168
Dy TCATGCGCATCCCGCAGCCG 65744
Qy 2168 yAlaSerCysProValLeuAsp-----LeuArgArgProProSerAspLeuTy 2184
Dy AGGCACAGCAAGCAACACAGACTGATTGAACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 65685
Qy 2184 rLeuProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLy 2204
Dy -----CGCTCCG 65630
Qy 2204 sArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluPr 2224
Dy CCCTCCACTCAGCAGCAGCG 65592
Qy 2224 oValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLe 2244
Dy -----CG 65547
Qy 2244 uLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLys----- 2260
Dy ACAGCGCTCGAAAAACATCCCGAAACCGCAACTCTCAGACCGCGCGCGCGCGCGCGCGCG 65487
Qy 2261 -----SerProGlyAsnThrSerGlnProProAlaPhePheSe 2273
Dy CAAGTGAATCGACAGAAATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 65427
Qy 2273 rLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLysLy 2293

| | | | | | | | |
|----|-------|---|-------|----|-------|---|-------|
| Qy | 626 | yLeuLeuGluHisGly-----ArgAsnTrpSerAlaIleAlaAraMetValG1 | 642 | Qy | 929 | AlaGluGlyGlyAspLysAsnArg | 936 |
| Db | 71279 | GTGTGTTGAAGCGAGCATGACCTGGAAAGGGGGTGGCGGCGAGGGAGCGGTTCGGGGTT | 71220 | Db | 70316 | CAGAACACCGGCCAGACGAGCGCCATCGGAATGTTGTACGACGACCGGGGCCCTTCGAG | 70257 |
| Qy | 642 | ySerLysThrValSerGlnCysAlysAsnPheTyrPhe-AsnTyrLysLysArgGlnAsnL | 662 | Qy | 937 | -----LeuLeuSerProArgProSerLeuLeuThrProThr----- | 948 |
| Db | 71219 | GACCGCTTGACACAGTCGTTCGAAAGGGATGTCTGGTGGCGTAGGCGTGTGAGTTCGG- | 71161 | Db | 70256 | TTCTCTAAGGAACCAACAGAGCGTGTCTGACGCAACGACACGCGTC-----CCGACCCGT | 70203 |
| Qy | 662 | euAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaAraArgL | 682 | Qy | 949 | -----GlyAspProArgAlaAsnAlaSerProGlnLysPro | 960 |
| Db | 71160 | -----TGCCCGCGAGCGCGCGAGAGTTCGCGCAACGTGGGTGCCCGGA | 71115 | Db | 70202 | CCGTCCTCGTCCACCAGAGGACCCGCGCACCCCAAAACCCGCGCTCCCGGAGCAACCG | 70143 |
| Qy | 682 | ysLysLysLysAlaProAlaAlaAser-GluGluAlaAlaPheProProValValGlu | 701 | Qy | 961 | LeuAspLeu-----LysGlnLeuLysGlnArgAlaAlaAlaIlePro----- | 974 |
| Db | 71114 | GAGTCGGTCCGAGAGGAGGTGTTCAGAGAAGCC----- | 71076 | Db | 70142 | CTCAGACGGCTCACCGTCGGCGCTCGAACAAACGCC-GCACACCCGCTCAACACCCA | 70084 |
| Qy | 702 | AspGlu-----GluMetGluAlaSerGlyValSerGlyAsnGluGluMet | 717 | Qy | 975 | ---ProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLys | 993 |
| Db | 71075 | GATGAGATCTTCGTGGCTGTCGTGGTGGCGCGCATCGGGTGCCTGGGATGTC | 71016 | Db | 70083 | ACACCGCAGCAGCAGCGGAATCAGACGAGTCGCCAGAAAGAGAAATGACCACCACTCGA | 70024 |
| Qy | 718 | ValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSer | 737 | Qy | 994 | -----ProAlaPro-----ProAlaProProProGlnAnLeu | 1005 |
| Db | 71015 | GTCCCGCGCGCCGAGCCG-----GGAGAGGAGCGC | 70986 | Db | 70023 | AGAGTCATCGTCCACCCCGCTCAACACCCAGCAGCTCCGCGAACACGACACACACA | 69964 |
| Qy | 738 | GlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThr---- | 756 | Qy | 1006 | GlnProGluSerAspAlaProGlnGlnProGlySer-----SerProArgGly | 1021 |
| Db | 70985 | GGCAGCGCGCCTGCAGCACCATTGAAGACCGTGACACCTCGGCACGCGCGAGTTCGC | 70926 | Db | 69963 | ACACTCTCCCGCGGGAACAGGCGCGGTAGCCGTCGCGCTCCACCTCCCGGCA | 69904 |
| Qy | 757 | GluAlaAlaLysAspThrGlyGlnAsnGlyPro---LysProProAlaThrLeuGlyAla | 775 | Qy | 1022 | LysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLys | 1041 |
| Db | 70925 | GAGGTCGGTGGAGCGTGGCGGGATGGTGAATCGATCGTCCCGCGGTAGTGCG | 70866 | Db | 69903 | CCGGAAGCGCGCGCTCCACTTCCCGTTCGCGGTCAACGGGATCTCATCCAGAACCA | 69844 |
| Qy | 776 | AspGlyProPro---ProGlyPro----- | 782 | Qy | 1042 | LeuProGlyAspProProCysTyrTrpThrSerGlyLeuProPhePro----- | 1056 |
| Db | 70865 | GACGGCGCGCGCGCGTCCGTGGGAGCGCAACTCGGCTGGGAGTGACGTAGTTG | 70806 | Db | 69843 | CCACCGCGAGCGCACCATGTATCGCAGGAGAGCGCTCAGCCACCAACACGAGATCGCG | 69784 |
| Qy | 783 | -----ProThr-----ProThr-----ProProArgArgT | 789 | Qy | 1057 | -----ValProProArgGluValIle | 1063 |
| Db | 70805 | GGTGGCCAGTAGCGAGCTGGGCGGACAGCTCGCTCGATCGTGTCTTCCTGGCGGAG | 70746 | Db | 69783 | AACGAAGACCATCCCAACACACCCCGCAGTCACTAAGCACCAACCCGCTGTATCAT | 69724 |
| Qy | 789 | hrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProp | 809 | Qy | 1064 | LysAlaSerProHisAlaProAsp-----ProSerAlaPheSerTyr | 1077 |
| Db | 70745 | TACGTGGCGCTGCCAGCGCGCATATCCGCGTACTGTCACGAGAGGGCGCGCAGCCGA | 70686 | Db | 69723 | CGCGCGGTCTCTACGACACCAACAACCCGACCGCACCATCAGGATGCGCGCAACAG | 69664 |
| Qy | 809 | roProAlaProProSerProSerAlaProPro-ProValValProLysGluGluLysGlu | 828 | Qy | 1078 | AlaProProGlyHisProLeuLeuGlyLeuHisAspThrAla----- | 1092 |
| Db | 70685 | GGCGGCTCTCGCGCGCTGGCGTGTAGCCCGTGGAGAGATCGCGCGAGCGGCCCAT | 70626 | Db | 69663 | CACCTCTCACCTCACCACTCCACACGGAACACCGAGCTTCACCTGTATCATCCGCAC | 69604 |
| Qy | 829 | GluGluThrAlaAlaAlaProProValGluGluGlyGluGlnLysProProAlaAla | 848 | Qy | 1093 | -----ArgProValLeuProArgProPro---ThrIleSerAsnProProProLeuIle | 1109 |
| Db | 70625 | CGACAGCGGTCCCGCGCATGTGTGTCAGCAGCACCAACGAGAGGAT----- | 70581 | Db | 69603 | GGCCAAACGAACCTAAGCGCGCGTCCACCCCGCACCGCACCAATCCCGCG-----TGC | 69550 |
| Qy | 849 | GluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGlu | 868 | Qy | 1110 | SerSerAlaLysHisProSerSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMet | 1129 |
| Db | 70580 | -----GTGGTGGTTCGTGGTGAGCGGAA-----GACTGTGGCGCG | 70545 | Db | 69549 | GATACATAGCTCACCCGCGCA----- | 69529 |
| Qy | 869 | GluAlaGlu-----GluGlyProAlaLysGlyLysAspAlaGluAlaAla | 883 | Qy | 1130 | SerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThr | 1149 |
| Db | 70544 | GACCGGAGATCGCGCGGAGTTCGAAGGGCGCGCGCGCGC-----TTGTC | 70497 | Db | 69528 | -----CACCGAACGACACGCCCAACACCGCTCCCGCTGACGCCACACCGCC | 69481 |
| Qy | 884 | GluAlaThrAlaGluGly---AlaLeuLysAlaGluLysLysGluGlySerGlyArg | 902 | Qy | 1150 | MetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGln | 1169 |
| Db | 70496 | GAGCGCTCTGAAGGGCGTCGGCACCGATGTCCAAAGAGGAGACCGGGAGCGGTGCG | 70437 | Db | 69480 | CGAGATAGCCCGCGCCA-----GACCCA----- | 69457 |
| Qy | 903 | AlaThrThrAlaLysSerSerGlyAlaPro-----GlnAspSerAsp | 916 | Qy | 1170 | GluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAla | 1189 |
| Db | 70436 | AGGGTCGTCAACGGCGGAGATGTGTCCAGCGACGTACCCGATCTCTCGCGATGAGGGT | 70377 | Db | 69456 | -----CACCCGCCCATCAAACTCACCCACACACCGACCGCGGACACCCCGCAGCG | 69406 |
| Qy | 917 | SerSerAlaThrCysSerAlaAspGluValAspGlu----- | 928 | Qy | 1190 | -----GlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValPro | 1204 |
| Db | 70376 | CGCGAGCTTTCATGGCGCGCATGACATCAATGAGCGCTGGTGCAGCGCTTCGAGTTC | 70317 | Db | 69405 | CCGCATCCAGAACAAACATCCGCTGACCC-GCGAAAGGAGCGCCCGATCGGCACAGCCCA | 69347 |
| Qy | | | | Qy | 1205 | GlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThr | 1224 |

| | | | | | | | | |
|----|-------|---|----------------------------|----|-------|--|---|-----|
| Db | 73442 | TTCCAGTTCTCAAGGAACACACAGACGCTGCTGAGCGCAACGACGCTCCCGACCCGT | 73383 | Db | 72419 | CAGGGGCTTCACCGCCACGACGCGGTGACGCGCGCGCAGAGGTGTGCCCGCGTCCGCGGAG | 72360 | |
| Qy | 74 | -----ProGlyAsnGluArgSerGlnGluLeuHisLeuArgPro----- | 86 | Qy | 370 | HisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu----- | Asn 384 | |
| Db | 73382 | CCGCTCCGTCGCCACGACGACGACCGCGCACCCCAACCGACCGCTCCGCGAGCAACCG | 73323 | Db | 72359 | CATGCGCTCCACCGACGCTCGGTGTCGCTGAGCATGTTGACCGCTGCTCGCACCGT | 72300 | |
| Qy | 87 | -----GluSerHisSerTyrLeuProGluLeu 95 | | Qy | 385 | LeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGln | 404 | |
| Db | 73322 | CTCCAAACGGCTCACCGTCGGCGCTCGAAACAACGCGCGCACACCGCTCAACACCCAA | 73263 | Db | 72299 | CTGGCGGCGCAGCGCGGTCGCGGACCGCTCTCCCGCTCCGCGAGACACGCGTGGCGCC | 72240 | |
| Qy | 96 | GlyLysSerGluMetGluPheIleGluSerLysArgProArgGluLeuLeuLeuProAsp | 115 | Qy | 405 | GlnArgIleLysPheIleAsnMetAsnMetAsnMetLeuMetAlaAspProMetLysValTyrLys | 424 | |
| Db | 73262 | CACCGCAGCAGACACGGAATACAGACGAGTCGCGAGAAGAATACCAACCCCAACCGAA | 73203 | Db | 72239 | GCTCAGTAGCGGAGCAAGAGCTCGAGGCGCGCGATGTGAAAGCCACCGTGTGTACGGA | 72180 | |
| Qy | 116 | ProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeu | 135 | Qy | 425 | AspArgGlnValMetAsnMetTyrSerGlu----- | GlnGluLysGlu 438 | |
| Db | 73202 | GAATCAT-CGTCCACCCCGCTCAACACCGACGACCTCCGGAACAGACACACACA | 73144 | Db | 72179 | CAGCAGCGGTGCGCCACCGCGGAGCCGCTGGGCACCATGTCCGCGCAGGAAGTTTAC | 72120 | |
| Qy | 136 | ThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHis | 155 | Qy | 439 | ThrPheArgGluLysPhe----- | MetGln 446 | |
| Db | 73143 | ACACCTCTCTCCGCG-----GCGAAACAGCGCGCGGTAGCGCGCGCTCCACCC | 73090 | Db | 72119 | CAGGCGCGCGACGCTACGACGACGCTTGGGCTGGCGGTGATCCCGAGGTGTAGAT | 72060 | |
| Qy | 156 | ThrAspProGluLeuGluLeuValProProArg-LeuSerLysGlu----- | GluLeuIleG1 174 | Qy | 447 | HisProLysAsnPheGlyLeuIle----- | AlaSerPheLeuLeuGluArgLysThrValAla 464 | |
| Db | 73089 | CCGCGCACCGAAGCGCGCGCTCCACCTTCCGCGCTCAACGCGGATCTCATCCA | 73030 | Db | 72059 | CACGTACGCGCGCGCGCGGCTTCCCGCGCGCGGCGGAGCGTGGCGGTGTCGTG | 72000 | |
| Qy | 174 | naanMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLys | 194 | Qy | 465 | GluCysVal----- | LeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyr 479 | |
| Db | 73029 | GAACACACCGCGCG-ACGCGACCATGTACGCGGCGAGCGCTCAGCAACGCGCGGATC | 72971 | Db | 71999 | GAGCTCGCGGTCACTGTCTCGGAGTCTGAGTACCATCATGGTCTCCCGCGGTATCGGGC | 71940 | |
| Qy | 194 | sLysLysGln----- | GlnGlnLeuGluGluG1 203 | Qy | 480 | LysSerLeuValArg---ArgSerTyrArgArgArgGlyLysSerGlnGln----- | 495 | |
| Db | 72970 | ACCGCAGGAGCAGTCCACCGGTGTAACACCGGTGATACGCAACCAACCGGTGATCA | 72911 | Db | 71939 | GCTGATGTCTCGTGTGAGGACCAACCGCGCGCGGTCTATCCAGCATGTAGTGTAG | 71880 | |
| Qy | 203 | uAlaAlaLysProGluProGluLysProValSerPro----- | 216 | Qy | 496 | ----- | GlnGlnGlnGlnGlnGlnGln----- | 505 |
| Db | 72910 | CCGCGCGGTCTCAGCAGCACAACACCGCGCAGCAGCAGCAGTCCCGCGCAACA | 72851 | Db | 71879 | CGCTCGCGCGGAGCCGCGGATCCAGCGCGAGATCTCCGCGACCTGTTTCTGTAGGCG | 71820 | |
| Qy | 217 | ---ProPro-IleGluSerLysHisArgSerLeu----- | ValGlnIleIleTyrAspG 233 | Qy | 506 | GlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLys | 525 | |
| Db | 72850 | GCACCTTCCACTCACCACCACTCCACACGGAACCGGACCTTACCTGATCATCGCA | 72791 | Db | 71819 | CAGCAGGCGCACGAGGAGCTCTTCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGG | 71760 | |
| Qy | 233 | luAsnArgLysLysAlaGluAlaAlaHis----- | 242 | Qy | 526 | GluLysGluAlaGluLysGluGlu----- | 533 | |
| Db | 72790 | CGACCAACAACTCAAGCGCGCGCTCCACCGCGCACCGCACAGATCCCGCGTACGCTAC | 72731 | Db | 71759 | GGCGCGCGCGCGCGCGCGAGCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG | 71700 | |
| Qy | 243 | -----ArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnG | 259 | Qy | 534 | ---GluLysProGluValGluAsnAspLysGluAsp----- | 544 | |
| Db | 72730 | ATACGCTCACCGCGCACGGAAGGGGCGAGCGCACGAA----- | CGCTCCGCGTC 72680 | Db | 71699 | GTACGTGACCGGATGTTCGCGAGACAGAGCGACATCGTGGCGGTGGCGCGCACCTG | 71640 | |
| Qy | 259 | ln-----ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlam | 276 | Qy | 545 | LeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu----- | 560 | |
| Db | 72679 | AGACCGCAGCCCGGA-GATAGCCCGCGGAGCGCGTCCCGCGGATGTACAGTCCGCC | 72621 | Db | 71639 | CTCTCGAACAGCTGGGCGGAGGAGTCCCGGGGCGGAGCGGCGGCGGCGGCGGCGG | 71580 | |
| Qy | 276 | etArgLysLysLeuIleLeuTyrPheLys---ArgArgAsnHisAlaArgLys----- | G 293 | Qy | 561 | ---LysGluAlaValAla-SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys | 579 | |
| Db | 72620 | TGCGACCGCGGTGGCAGCAGACACAAAGCAGCGTCGAGCACATACAGCGAGTGTGAG | 72561 | Db | 71579 | GTGACAGTGTGTGCGCTCGCGCGCGCGGAGCAGCTCGACGCTCTCGCACGTTGCTC | 71520 | |
| Qy | 293 | lnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysV | 313 | Qy | 579 | sgly----- | ArgIleThrArgSerMetAlaAsnGluAlaAsnSer----- | 592 |
| Db | 72560 | GATGG----- | GTGTGCGGATCGAAGG 72540 | Db | 71519 | CGGGGCTCGCGCAAAAGCAGGAGACATGCGGAACCGGTTCGCGGAGCTGCGCGCAGGT | 71460 | |
| Qy | 313 | alGluArgIleGluAsnAsnProArgArg----- | AlaLysGluSerLysVal-Arg 330 | Qy | 593 | ----- | GluGluAl 595 | |
| Db | 72539 | GAGAGCGCTCTCCCGCGCGCTGCGCGCGCTGGTGAGACAGATGCTCTCCGTCGG | 72480 | Db | 71459 | GTCTCGTCCACAGCTCGCGCGGTACTCGAAGACACTTCTGATGCCCTGGTGGTGGCG | 71400 | |
| Qy | 331 | GluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMet | 350 | Qy | 595 | alleThrProGlnGlnSer----- | AlaGluLeuAlaSe 606 | |
| Db | 72479 | GCGGTACATGTCTGTGACGAGCAGCGCGCTCGTGTGAGCGCTCGCGCAACCGCGGCT | 72420 | Db | 71399 | CGTCTCACCGCGGATTCGTTGCAATACAGGCGCGAGATCGAATTCGCCACCGCGGCTC | 71340 | |
| Qy | 351 | GlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArg----- | SerGlu 369 | Qy | 606 | rMetGluLeuAsnGluSerSerArgTyrThrGluGluGluMetGluThrAlaLysLysG1 | 626 | |
| Db | | | | Db | 71339 | GATGTCTCGCGGAGGCGGAGCGGCTTCGCGCATGGGCGGCGGCGGCTCTCTCTCGCG | 71280 | |

Db 3254 CGTTGGCGCGCGCGCATCGCCACCA-----TTGCGCGCGCGCCACCA 3301
Qy 1560 ValThrMetArg-----GluProThrProArgLeuGlnGluGlySerLeuSer 1576
Db 3302 TCGCCACCGCGCGCGCATTCGCGCGTCCGTTGCC----- 3343
Qy 1577 SerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluLeuAlaLysSerPro 1596
Db 3344 -----CCGCGCGCGCGCGCGTGGCTTCGCGCACCG 3376
Qy 1597 HisSerThrValProGluHisHisProHisProHisProHisProHisProHis 1616
Db 3377 TTGCCACCGTACACCGTTCGCGCGTCCG-----GGCCGCGTGG----- 3418
Qy 1617 GlyValSerGlyValAspLeuTyArgSerHisLeuProLeuAlaPheAspProThrSer 1636
Db 3419 TCCTGGGTCCACACCGT-----GCCGCGCGGTGCGTTCGCGCGCCCAAG 3475
Qy 1637 IlePro-ArgGlyIleProLeuAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1656
Db 3476 ---CCGCGGAGGCTCCGCGCGCGGTACCGCGGTACCGCGTCCCGCGCATCA-CT 3531
Qy 1656 aProAsnProThrTyProHisLeuTyPro-----ProTyLeuIleArgG 1672
Db 3532 GCCATGACCGCGTCAACCGCGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGCAA 3590
Qy 1672 yTyProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyIleThrSe 1692
Db 3591 CGCGCGGACACCGTGCT----- 3609
Qy 1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgG 1712
Db 3610 -----GCCGACCTGACCGCGG 3626
Qy 1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyAlaAlaGlyProArgGlyIleI 1732
Db 3627 GTGCGCGCGCGCGCGCGCATC-----ACGCGCGCGCGC----- 3663
Qy 1732 eAspLeuSerGlnValProHisLeuProValLeuVal-ProProThrProGlyThrProA 1752
Db 3664 -----GCCGTACCGCCCAACCGCGCGCACCGCG 3692
Qy 1752 laThrAlaMetAspArgLeuAlaTyLeuProThrAlaProGlnProPheSerSerArgH 1772
Db 3693 CGTCGCG-----CCTTCCCGCGCGCGCTCCCAACGCGCTCAG 3731
Qy 1772 isSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrHis 1792
Db 3732 CGTGTGCGCGCGGAGCGCGC-----TTGCGCGCGCGCGC-ACCACTA 3775
Qy 1792 erSerSerGluArgGluArgAspArgGluArgAspArgAspArgGluArgGlu 1812
Db 3776 GTTCGCG-----ACCACCGCACCGCGCGC 3802
Qy 1812 ysserIleLeuThrSerThrThrValGluHisAlaProIleTTrpArgProGlyThrG 1832
Db 3803 CGGCGCTTGCACCGCGTCCACGTCGCGCGCACCGACCC----- 3843
Qy 1832 luGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1852
Db 3844 -----ACCGCGCGCACCGGTCCGCG-----GGCACCGCGCGCGCGC- 3888
Qy 1852 laSerHis-----SerHisAlaHisG 1859
Db 3889 CCGCGATCGCTCGACCGCGGTTCGCGCGCACTACTACCGCGCACTCCCGCACCAACC 3948
Qy 1859 lniHisSerProIleSerProArg-----ThrGlnAspAlaLeuGlnGlnArg---ProS 1876
Db 3949 GTACCGCGCGCACCG 4008
Qy 1876 erValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProT 1896
Db 4009 GCTTCTTCGCGCGGTTCGCGCGGAGGGTGCAG-TCGCGG-----CCGTCTCTCACCTA 4061

Qy 1896 hrValLeuArgSerThrSerSerProValArgProAlaAlaThrPheProProA 1916
Db 4062 CGTTC-----GCGCATTCGCACCGCGCGCGCTACCAACCGT 4100
Qy 1916 laThrHisCysProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProV 1936
Db 4101 CGCGCGCGCGCGCGTCAACCGACCAACCGCGTTCACCGCGCACCGCGCGCTTCGCGC 4160
Qy 1936 alLeuLeuPro-----LysGluAlaProArgValAlaArgProGlu----- 1949
Db 4161 CGTTCGCGCGCACCGCGTGGCGTTCGCGCGCTTACCGCGCTTCGCGCGTTCGCCACAC 4220
Qy 1950 -----ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysPheProAlaA 1966
Db 4221 TGTGACGCGTAGCGCGTCCGCGGTACACCTCGGTGGCGCGCGCGCGCGCG- 4279
Qy 1966 rgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValP 1986
Db 4280 -----CCACCTTCCACCGCGCGTGGGTACACCATCGCGCGCGCGCGC 4325
Qy 1986 roProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisH 2006
Db 4326 CACCG-----TCACCATCTTGAAAGCATGTGCGCGCGCGCACCGTCCCGCGTTAC 4379
Qy 2006 isAlaSerProAspProAlaProAlaSerAlaSerAspProHisArgGluLysT 2026
Db 4380 CGCGCGCG-----CCACCGCGCGCGCGCACCGTCCCGCGCA----- 4420
Qy 2026 hrGlnSerLysProPheSerIleGlnGluLeuLeuArgSerLeuGlyTyHisGlyS 2046
Db 4421 -----ACACCTCG-----CTGCGCGCGCG----- 4441
Qy 2046 erSerTySerProGluGlyValGluProValSerProValSerProSerLeuThrH 2066
Db 4442 -----CCCGAGGTGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4490
Qy 2066 isAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyL 2086
Db 4491 GACCGCGCGTCCGCGCATCA-----CCACCGTTCGCGCGTTCAGATCAGACAGCAT 4544
Qy 2086 euArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisL 2106
Db 4545 TGAACCGTGAAGTTCGTTACCA-----CCGCGCGCGCGCGCGCGCGCGCGC 4589
Qy 2106 euArgProLeuProGluSerGlnProSer----- 2115
Db 4590 TGCAGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4649
Qy 2116 -----SerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValV 2134
Db 4650 CACCGCGCGCGCGTTCATCTCCATCTCCTACCGCGGTTCGCGATGAACCCAGGATCG 4709
Qy 2134 hrLeu-----AlaGlnHisIleSerGluValIleThrGlnAspTyThrArgHisP 2152
Db 4710 CGGTATATCGCATAGCGCGCACCGCGCGTCCGTACCGCGCATGACCGCGCGCGCGC 4769
Qy 2152 roGlnGlnLeuSerAlaProLeuPro-----AlaProLeuTySerPheProGlyAlaS 2170
Db 4770 GTTACCGACCAACCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4829
Qy 2170 er-----CysProValLeuAspLeuArgArgProProSerAspLeuTyL 2185
Db 4830 CGGTGCGTGGCGCGCATTCGCCACCTTTCGCGCGGTTCGCGCGTTCGCGCGCGCGC 4889
Qy 2185 euPro-ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyS 2204
Db 4890 TGCAGCGCGCGCACCATTCG-CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4948
Qy 2205 ArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluPro 2224
Db 4949 GCTCCGC 4972


```
QY 2306 IleSerGlnProGlyThrGluIlePhe-AsnMetProAlaIleThrGlyThrGlyLeuMe 2325
Db 105226 -----CGGCGCCATTATTCCGGTAGCCCGGCACCGCGGCACCGCGCGCG 105273
QY 2325 tThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaI 2345
Db 105274 CACCGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105328
QY 2345 eIleArgLysAlaLeuMetGlyLysTyArgGlnTrpGluGluSerProProLeuSerAl 2365
Db 105329 -----GAGAACCCACCTCCCGCGGCAC 105351
QY 2365 aAsnAlaPheAsnProLeu---AsnLaserAlaSerLeuProAlaAlaMetProIleTh 2384
Db 105352 CGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105411
QY 2384 rAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLys 2404
Db 105412 CGCGCGG-----ACCGTCGTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 105453
QY 2404 sValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGl 2424
Db 105454 CTTGACCGCGCGTAGCACCGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105513
QY 2424 y-----AspArgProSerValSerSerValHisSerGluGlyAspCysAsnArgAr 2442
Db 105514 CCCCATGACCAACACCGCGCGCGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105570
QY 2442 gThrProLeuThrAsnArgVal-TrpGluAspArgProSerSerAlaGlySerThrProp 2462
Db 105571 CACCGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105630
QY 2462 hePro-----TyrAsnProLeuI 2468
Db 105631 CGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105690
QY 2468 leMetArgLeuGlnAlaGlyValMetAlaSerProProProProProGlyLeuProAlaGlyS 2488
Db 105691 TCCCGCGCGTAGCTCCGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105750
QY 2488 erGlyProLeuAlaGlyPro 2494
Db 105751 ATAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105770

RESULT 38
AAI99682_28
Continuation (29 of 45) of AAI99682 from base 2800001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
```

```
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Alignment Scores:
Pred. No.: 4,26e-11 Length: 110000
Score: 664.00 Matches: 481
Percent Similarity: 33.79% Conservative: 182
Best Local Similarity: 24.52% Mismatches: 774
Query Match: 5.02% Indels: 537
DB: 4 Gaps: 86

US-09-522-753-5 (1-2517) x AAI99682_28 (1-110000)
QY 685 LysAlaProAlaAlaAlaSerGluGluAla-----AlaPheProValVal 700
Db 1034 AAGTCGCGCGTAGCACCGATCGCGGAGTACACCGCGCGCGTGGGGTTTCGAGGCTAGG 1093
QY 701 GluAspGluGluMetGluAlaSerGlyVal-----SerGly 712
Db 1094 AATGATGACGATGGGCGCGCGCTCGCGCGCTTCCGCGTAACCCCAACCATCGCGAA 1153
QY 713 AsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGly---AsnGluVal 731
Db 1154 AACGAGCCTAGCGTCCGCCGCGCGCGAGAGCAGCATCGCGTGGCGCGCAACGACGAG 1213
QY 732 ProArg-----GlyGluCysSerGlyProAlaThrValAsnAsnSer 745
Db 1214 AACGATCCGGATTCTCTGACCATGTTGGTGTCTTGCTACGTGAGTTA-----1264
QY 746 SerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn 765
Db 1265 ACGGAGATGGAGGCGCGCTTCGCG-----GAACCGCGGAGCGAGGTCAAGCCACCGCA 1291
QY 766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro 785
Db 1292 ---CCTTACCGCGCG-----1333
QY 786 ProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla 805
Db 1334 TTGCCGGAGCCACCGACTCGCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1393
QY 806 ProThrPro-----ProProAlaProProSerProSerAlaProProValPro 823
Db 1394 CCGCGTCCGCGCGTACGTCGCGCGCGCGTCCCGCGCGTCCCGCGCGCGCGCGCGCGCG 1452
QY 824 LysGluGluLysGluGluThrAlaAlaAlaProPro-----ValGluGlu 839
Db 1453 AGCGGAGGGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1512
QY 840 GlyGluGluGlnLysProProAlaAlaGluGluAlaValAspThrGlyLysAlaGlu 859
Db 1513 GGCGGCGCGCACCGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1569
```



```
QY 945 LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys 964
Db 101820 CCAGCGTCCCGCCGCGCGTCCAGCGCGTCCCGCCCAACGCT-----CGA 101870
QY 965 GlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluPro 984
Db 101871 TTGCCACCTCGCGCGCGTCCGAGCCGCG-----GCCCG 101909
QY 985 ProArgGluAspAlaAlaProThrLysProAlaProProAlaProProGlnAsn 1004
Db 101910 CCGTTTCGCGCGCGCCGCGCCACCGCGGTAGCACCGTCCGCGCGCGTCCGCGCGTCCG 101969
QY 1005 -----LeuGlnProGluSerAspAla 1011
Db 101970 CGGCCCGCTGTTTCGTTAATGTCAAGCGGTCAACGCGTTACCGCGCGACCCACCGCC 102029
QY 1012 ProGlnGlnProGlySerSerProArgGlyLysSerArg-----SerProAlaPro 1028
Db 102030 CCGCTTAGGCTCCGCGCGCGCGCGTCCACCGTCCGCGTCTGAGTTCCGCGCGCGCA 102089
QY 1029 ProAla-----AspLysGluAlaPhe 1035
Db 102090 CCGGCCCGCGTCCGCTCCGCGCCCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCG 102149
QY 1036 AlaAlaGluAlaGlnLysLeuProGlyAspPro-----ProCysTrpThrSer 1051
Db 102150 AGTCGGAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 102206
QY 1052 GlyLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAsp 1071
Db 102207 -----CCGCTTCGCGCGCGCGCGTTCATCCGCGCGTCCGCGCGCGCGCGCGCG 102254
QY 1072 ProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThr 1091
Db 102255 -----TGTCCACCAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 102287
QY 1092 AlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSe 1111
Db 102288 -----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 102329
QY 1111 rAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVa 1131
Db 102330 CCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 102360
QY 1131 lGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGl 1151
Db 102361 -----CACAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 102412
QY 1151 yLeuPro---LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGl 1170
Db 102413 CCCTCCGAGGCGACCGTCCGCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCG 102458
QY 1170 uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGl 1190
Db 102459 -CGCGCGTACCGGATGCGCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 102505
QY 1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy 1210
Db 102506 ACCGGCCCC-GCCATCACCGCGAGGTTCCGCGCGCGTCCGCGCGCGCGCGCGCGCG 102564
QY 1210 sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
Db 102565 CGGCTCCGCGCGTCCCGAATAGCCGCGGTTCGCGCGTCCGCGCGTCCGCGCGCGCGCG 102624
QY 1230 rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAs 1250
Db 102625 CGCGCGCGCGCGCGCG-----CTTGGCACCCTTGC----- 102655
QY 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy 1270
Db 102656 -GCCGAGCGCGCGTCCGCGCGCGTCCAGTCCGCGCGTTCGCGCGTTCGCGCGTCCACCGCG 102712
```

```
QY 1270 rGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe 1290
Db 102713 -----GGCCCGCGAGGTGCGGTTCGCGTTCACCCCGACAGACCGCGCGCGCGCG 102768
QY 1290 rLysGluAspGlyArgSerSerClyProProHisGluThrAlaAlaProLysArgTh 1310
Db 102769 C-----GGGGTCCACCGCGCTTACCGCCCGCTC 102795
QY 1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 102796 CGCCGCGCGCGCGCGCGGTATACCAATCGCGCGT---GCCCGCGCGCGCGCGCAACGC 102852
QY 1330 uMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisH 1350
Db 102853 CGCGCGCGCGCGCGA-GTCCGCGGTCCGCGCGCGCGCGCGCGCGTCCGCGTCCGCGA-- 102909
QY 1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTy 1370
Db 102910 -----CCGTCCCGAAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 102941
QY 1370 rLeuArgArgGluAlaLysLeuLysArgGluGlyThrProProProProProProSe 1390
Db 102942 CCGTCCCGCGCATGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103000
QY 1390 rArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
Db 103001 CCACCGGTGCGCGCGCATACTGTGATA---CGTTGGCGCGCGCGT----- 103043
QY 1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
Db 103044 -----GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103051
QY 1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly-- 1449
Db 103052 CCCACCGCTAGCGCGCGCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103107
QY 1450 -SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db 103108 CGCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103141
QY 1469 sLysHisAspValArgSerLeuIleGlySerProGlyArgThr-PheProProValHisP 1489
Db 103142 -----GGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103176
QY 1489 rLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuL 1509
Db 103177 CC----- 103178
QY 1509 ySerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaProValI 1529
Db 103179 --GCGTCCGCTGACCGCGCGTTCCTCCCGGTTCGCGCGCGCGCGCGCGCGCGCG 103229
QY 1529 leValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlap 1549
Db 103230 -----GAACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103281
QY 1549 rPheAlaGlyHisLeuProArgGlySerProValThrMetArg-----GluProT 1566
Db 103282 CA-----TTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103329
QY 1566 hrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db 103330 CGGCTCCGTTCGCG----- 103344
QY 1586 hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProH 1606
Db 103345 CGCGCGCGCGCGCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103404
QY 1606 isProIleSerProTyrGluHisLeuArgGlyValSerGlyValAspLeuTyrArgS 1626
Db 103405 TGCGG---GGCGCGTGG-----TCGCTGGCGTCCACACCGCGT---GCCG 103443
QY 1626 erHisIleProLeuAlaPheAspProThrSerIlePro-ArgGlyIleProLeuAspAla 1645
```



```
Db 99753 CCGCGATCGGTCCCGCTCGTCCCAATACCTCAGCGTGGGACGCCCGGAGCAACG 99812
Qy 353 gValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis---GluVa 372
Db 99813 CGACGATCAGAGCGCGCACCGCGTCGAGCAAGTCTCGCAGTTGTCCAGCACTACCAGCA 99872
Qy 372 lSerGluIleAlaAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLe 392
Db 99873 TGGCGGTGCGCCGAT-----ACGCGGCACAATAGTGTCCACCGTCCAGCGGCCCGGCT 99926
Qy 392 uAlaValIlePro-----ProMetLeuTyArgAlaAspGlnGlnArgIle----- 407
Db 99927 GATCCGGAACCCCAAAACCGCGCGCGCGGATCGGCACACGCGCGGTTCGTGTATCG 99986
Qy 408 -----LysPheIl 410
Db 99987 GCGCCAGTTGACATACCAACACCCCGTCGGATAAACCGTCGGCAACGGCGTCGCGACCT 100046
Qy 410 eAsnMetAsnGlyLeuMetAlaAspProMetLysValTyLysAspArgGlnValMetAs 430
Db 10047 GTGTCCGACGCGTGTCTTTCCGACCCCGCGCGACACCGGTAAAGGTGACCCACCGTTTGA 100106
Qy 430 nMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAs 450
Db 100107 CGTC-----CAGCAGCCACCGGACTTCGGCCACTTCGTGACGCGGCCACCA 100154
Qy 450 nPheGlyLeuIleAlaSerPheLeuGluArgLysThrVal----- 463
Db 100155 G-----CCGAGTGAAGTGGCGCGGAGACAGTGGCGCACCAACGACTTTCGCGG 100202
Qy 464 -----AlaGluCysValLeuTyTyTyLeuThrLysLysAsnGluAsnTyLys 480
Db 100203 TCCGACGCGCGGGAACCGTGTGTCAGATCAGGTTGACAGCTGACACACCGCTTCCG 100262
Qy 480 sSer-----LeuValArgSerTyArgArgArgGlyLysSerGl 494
Db 100263 GTCGGGCGAGTGTCCAGCGGTAGGTACCGAGTGTTCAGCAGCGGTCTTCGGCA 100322
Qy 494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
Db 100323 GCAGTCAAGCAACAGATCGTGTGTAGTTCCCGACAACACGCTGTGGCCCCCGTGGGCA 100382
Qy 514 sSerSerGln----- 517
Db 100383 GCTCGCGCAGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 100442
Qy 518 -GluGluLysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysPro-- 536
Db 100443 ACACACCTCCCGGTGTGTCATGCGATGCGCAGCGCGATCGGTCCAGCGCGCGCGCT 100502
Qy 537 -GluValGluAsnAspLysGluAspLeuLysGluLysThrAspAspThrSerGlyGl 556
Db 100503 GCAAGCCAGGCGCGCACCCACCGCGTC-----GGATCGCGCGGCGA 100544
Qy 556 uAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThr-----Al 572
Db 100545 ACGCCACCAAGAGTGTGCTTCCCTGCTTCCAGCGGCAACACCCCGGTGTCTCG 100604
Qy 572 aAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSe 592
Db 100605 GAACCAATTGCTCAGCTTC--GGTCCAGTTTGGCGATCGCGCTGTCAAGTGTAGAC 100663
Qy 592 r-----GluGluAlaIleThrProGlnGlnSerAl 602
Db 100664 CCGCGCAGTGGGTCCGCGCTCGATATCGGCACAGACAGTCACCGTCCCGTCCGT 100723
Qy 602 aGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu----- 618
Db 100724 ACAAGCTCGCTCACACCATCTCGCTCCAGTCCACAGT--ACCAGTTCGACGCGCGGTG 100782
Qy 619 -----GluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
```



```
QY 2479 oProProGluLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAs 2499
Db 53891 GCGCGCGCGTCCGACCGCGTCCGTCCTCTACGGCC-----53930
QY 2499 pGluGluProGlyProLeuLeuCysSerGlnTyrGluThr 2512
Db 53931 ----CGACCGCGCACCGAGTACCTGTGGACGAGGATCTACC 53966

RESULT 37
Continuation (28 of 45) of AA199682 from base 2700001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682
WP Fragment Name Begin End
WP AA199682_00 1 110000
WP AA199682_01 100001 210000
WP AA199682_02 200001 310000
WP AA199682_03 300001 410000
WP AA199682_04 400001 510000
WP AA199682_05 500001 610000
WP AA199682_06 600001 710000
WP AA199682_07 700001 810000
WP AA199682_08 800001 910000
WP AA199682_09 900001 1010000
WP AA199682_10 1000001 1110000
WP AA199682_11 1100001 1210000
WP AA199682_12 1200001 1310000
WP AA199682_13 1300001 1410000
WP AA199682_14 1400001 1510000
WP AA199682_15 1500001 1610000
WP AA199682_16 1600001 1710000
WP AA199682_17 1700001 1810000
WP AA199682_18 1800001 1910000
WP AA199682_19 1900001 2010000
WP AA199682_20 200001 2110000
WP AA199682_21 2100001 2210000
WP AA199682_22 2200001 2310000
WP AA199682_23 2300001 2410000
WP AA199682_24 2400001 2510000
WP AA199682_25 2500001 2610000
WP AA199682_26 2600001 2710000
WP AA199682_27 2700001 2810000
WP AA199682_28 2800001 2910000
WP AA199682_29 2900001 3010000
WP AA199682_30 3000001 3110000
WP AA199682_31 3100001 3210000
WP AA199682_32 3200001 3310000
WP AA199682_33 3300001 3410000
WP AA199682_34 3400001 3510000
WP AA199682_35 3500001 3610000
WP AA199682_36 3600001 3710000
WP AA199682_37 3700001 3810000
WP AA199682_38 3800001 3910000
WP AA199682_39 3900001 4010000
WP AA199682_40 4000001 4110000
WP AA199682_41 4100001 4210000
WP AA199682_42 4200001 4310000
WP AA199682_43 4300001 4410000
WP AA199682_44 4400001 4411529

Alignment Scores:
Pred. No.: 3,97e-11 Length: 110000
Score: 665.00 Matches: 627
Percent Similarity: 33.04% Conservative: 296
Best Local Similarity: 22.44% Mismatches: 1111
Query Match: 5.03% Indels: 776
DB: 4 Gaps: 122

US-09-522-753-5 (1-2517) x AA199682_27 (1-110000)

QY 20 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 39
Db 98750 CCCCCGACGCGGTACGCGC-CGCGCCACGACGAGGCGCG-----ATC 98790
```

```
QY 40 LeuGluTyrGlnHisHisSerArg-----AspTyr 49
Db 98791 CTCGATGTCGCGGCTATAGCCAGCGCCTTCTGGGCATCATCTAGGCGGTGCGTGATGCC 98850
QY 50 AlaSerHisLeuSerProGlySerIle-----IleGlnProGlnArgArg 64
Db 98851 GGCACGCGCATCTATCAGGCGCGGTGCGGCCAGTGTCCGCGCATACAACCCAGGGT-CGG 98909
QY 65 ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeu 84
Db 98910 CCCCCCGCGATGTGTCATCGTGTGCGCAGCGCGCGGCGGAGACAGCCAGCCCTCTT 98969
QY 85 -----ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer 98
Db 98970 GCAGGCGGCGCGCGGCGCACCAACGGCTGCAGACATGATGCCAACAGCAGCGGTGC 99029
QY 99 GluMetGluPheIleGluSer-----LysArgPro-----ArgLeuGluLeu 112
Db 99030 CGGTATCGCATTCCTCGCGGTGAACGCGAAGCGCGCGGTAGGTGTGTCATCTCGA--- 99086
QY 113 LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer 132
Db 99087 -----GCTCGGCTGTTGAGCCGCGGCTTCATGCGCGCCACCGAGGGG 99131
QY 133 GluAspLeuThrLysAspArgSerLeuThrGly-----Lys 144
Db 99132 CGTCAAGCGCGGCGGCAACCGCGCGTAGTGTGCGGTGTCGCGCAGCAGCAGCATCGG 99191
QY 145 LeuGluProValSerProProSerProHisThrAsp---ProGluLeuGluLeuVal 163
Db 99192 CATCGCGCGATTTCACGCGCTTCTCCACGCGCATCTGGCGCACCGCTCTTAGCAGCGGT 99251
QY 164 ProPro-----ArgLeuSerLysGluLeuLeuIleGlnAsnMetAspArg 178
Db 99252 AGCGCGTTCGCGCGCTGCTGCTCGTCCGTCACACAGAGACTGTCTGCCAGCAGCGTGA 99311
QY 179 ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLysLysLysGlnGln 198
Db 99312 GCAGATCGCACCATCTGTAGCGCTGAACGTCACCGCGCGGCTGCGGTGGGACCGT 99371
QY 199 GlnLeuGluGluAlaLysProProGluProGluLysProValSerProProPro 218
Db 99372 CGAGATCAAAACCGCTCGGGAACCGCCAGTCGCCGAA-----ACAGCACCTGTCTCCG 99425
QY 219 IleGluSerLysHisArgSerLeuValGlnIleIle-TyrAspGluAsnArgLysLysAl 238
Db 99426 GTC-----CGTCAGCAGCGCATGTACCAGTCGACGGAAGCCCGCATCGTCT 99473
QY 238 aGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyAs 258
Db 99474 GCTGCGCGCGCACCGCAATACGCGATCCACCGTTCAGCAGCGGAA---CCGCTCATGCA 99530
QY 258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
Db 99531 AGCTGTGACGATTCGCGTACG-----ArgIleGluAsnProArg----- 99552
QY 278 sLysLeuIleLeuTyrPhenylsArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh 298
Db 99553 -----CGCCGCGCGCACCGCGGAGCC----- 99576
QY 298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu----- 314
Db 99577 -TGC---AAGTTCGATCGCGCGGGAATGCGTCGATCGGTGGCAGATCTCGGTACCA 99632
QY 315 -----ArgIleGluAsnProArg----- 321
Db 99633 GGGCGAGTTGTGCGCAGTATCTCGATTTCGCGCGCGCTCACGACGCGGTGCGTGA 99692
QY 322 -----ArgArgAlaLysGluSerLysValArgGluTyrTy 333
Db 99693 ACAATCGATCGCTCGCGTCCCGCGGGAACCCGCCAAATCTGTCCACCGGCCA 99752
QY 333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerAr 353
```

```
Db 52019 -----CGCCCGCGCACTGGGACGTGGCGCTGAAC 52047
Qy 1869 pAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysIleIleThrAl 1889
Db 52048 TTCTGCCCAACCGCCCA----- 52066
Qy 1889 aValGluProSerLysProThrValLeuArgSerThrSer-----Th 1903
Db 52067 -----CAAGCGCGCGCAACCGCGCGCAACCGCACCTCGCGAAGCGCGCTGTGCATC 52119
Qy 1903 rSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyG1 1923
Db 52120 GAGTGTTCGCGAGCGCGACCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 52179
Qy 1923 yThrLeuAspGlyValTyProThrLeuMetGluProValLeuLeuProLysGluAlaPr 1943
Db 52180 GACGG--CACGGTGTGCGCCT-----CTCCCTAGGCGAAGCTCG 52216
Qy 1943 oArgValAlaArgPro-----GluArgProArgAlaAspThrGlyHisAlaPheLe 1960
Db 52217 GCGACGACACCGTGTGCTCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52253
Qy 1960 uAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerG1 1980
Db 52254 -----CCACCGCGCACCGAACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52306
Qy 1980 uPro-----ArgProLeuValProProValSerGlyHisAlaThrIleAlaAr 1996
Db 52307 TCCTGGCGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52354
Qy 1996 gThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProAl 2016
Db 52355 CCCC CGCGCA-----CACCAAGAGAGCGACCGATGACCGACCGCGCGCGCGCG 52399
Qy 2016 aSerAlaSerAspPro-----HisArgGluLysThrGlnSerLysProPheSerIle-- 2033
Db 52400 GACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52459
Qy 2034 -----GlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerPr 2050
Db 52460 GCTCGCGCGCAACAGGAGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52515
Qy 2050 oGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLe 2070
Db 52516 AGAGGAGTG-----CCTGCGCGCACATCGACGAGTG 52548
Qy 2070 uProLysHisLeu-----GluGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
Db 52549 GACCGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52608
Qy 2088 -ProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuAr 2107
Db 52609 CGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52665
Qy 2107 gProLeuProGlu-----SerGlnProSerSerProLeuLeuGlnThrAlaPr 2124
Db 52666 CGGCGACACGCGCGGTGGAGTGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52725
Qy 2124 oGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleTh 2144
Db 52726 AGGCGGTGACCGCGGCTACGTCTTGTCTGCGTCTGCGCGCGCGCGCGCGCGCGCGCG 52785
Qy 2144 rGln-----AspTyThrArgHis----- 2150
Db 52786 TCGGCGTTCGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52845
Qy 2151 -----HisProGlnGlnLeuSerAlaProLeuProAlaPro-----Le 2163
Db 52846 GCACCTCTACCTCTGATCTGGAACCGGTCTCGTGTGCGCGCGCGCGCGCGCGCGCGCG 52905
Qy 2163 uTySerPhe-----ProGlyAlaSerCysProValLeuAspLeuArgProPro-- 2180
Db 52906 TCACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52964
```

```
Qy 2181 -----SerAspLeuTyLeuProProAspHisGlyAlaProAlaArgGlySerPr 2198
Db 52965 CCGTAGGCGCGCGCTCGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53015
Qy 2198 oHisSerGlu----- 2201
Db 53016 CCATGACCAACCCCATGACACCCCGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 53075
Qy 2202 -----GlyGlyLysArgSerProGluPr 2209
Db 53076 TCGCCACACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53135
Qy 2209 oAsnLysThrSerValLeuGly-----GlyGlyGluAspGlyIleGlu-----ProValSe 2226
Db 53136 TGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53195
Qy 2226 rProProGluGlyMetThrGluProGlyHisSerArg-SerAlaValTyProLeuLeuT 2246
Db 53196 CCGCTTCGTTCGGGCACTCACCGCGCACACGTACGCCACCGACCGCGTGTGCGCTCG 53251
Qy 2246 yzArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrS 2266
Db 53252 -----CCCGACGGG-----GGCGAGGCGCGACCCC----- 53275
Qy 2266 erGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMet-----ValL 2284
Db 53276 --CGTGGCGCGCGCTTCCTCGCGCACCTTGGCGCGCGCGCGCGCGCGCGCGCG 53333
Qy 2284 ySerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArg-----AsnGluP 2302
Db 53334 ACCCGACACGACGGCGCTGCACGCC-----GGCCACCGCGCGCGCGCGCGCGCG 53384
Qy 2302 roGlu-----TyAsnIleSerGlnProGlyThrGluIlePheAsnMetPro----- 2317
Db 53385 CCGAGCGGATCGCGCTTCGCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53444
Qy 2318 -----AlaIleThrGlyThr-----GlyLeuMetThrTyArgSerGlnAlaValG 2333
Db 53445 GCCTCCACGCTCCCTGTGGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53497
Qy 2333 lGlnLysHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyL 2353
Db 53498 -----CTCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53539
Qy 2353 ySTyrAspGlnTrpGluGlu-SerProLeuSerAlaAsnAlaPheAsn----- 2369
Db 53540 -----GACGTCTCCCGGAGTGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53594
Qy 2370 ---ProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGly 2388
Db 53595 TCGCCCGCTCGCGCTCGTCCGCGTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53654
Qy 2389 ArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArg 2408
Db 53655 CGCTCGAGC-----ACTACCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 53705
Qy 2409 ProSerSer-----ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAs 2425
Db 53706 CCTCCCTCTGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53765
Qy 2425 pArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLe 2445
Db 53766 CCGTCTCTGACCGGTCCCGCG-----AGCCGTCTCTCCACCTCG 53804
Qy 2445 uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPro----- 2461
Db 53805 AGGCGCGCGCGCG-ACCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53863
Qy 2462 -----PheProTyTrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPr 2479
Db 53864 GCGAACGACTCTCTGCTCGCG-----ACGTC 53890
```

QY 1300 ----ProPro-----HisGluThrAlaAlaProLeuArgThrTyrAspMetMetG1 1315
Db 50124 GCACCCACCTCGACACCGTCAACCGGTGGCTGACCGCTACGGAGC----- 50169
QY 1315 uGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaI1 1335
Db 50170 -----CGTCTGT 50177
QY 1335 eProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerI1 1355
Db 50178 TCGCGCGTTCGGCGTTCGCGCTCGACGGTTCGCGACGTCGTCGACGCGCTCGCGGAT 50237
QY 1355 eThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAl 1375
Db 50238 CCCCCGAGGGTACGTGCAACGGTCTGTCGCGCGCAC-----CGCCCTCGGCG 50285
QY 1375 aLysLeuLeuLysArgGlu-----GlyThrProProProProProProSerArgAs 1392
Db 50286 ATCACCCTTACACCGCCACCGACCCCGCCGACCGACCGACCCATCCCCCGCACACGAGA 50345
QY 1392 pLeu-----ThrGluAlaTyrLysThrGlnAlaLeu----- 1402
Db 50346 ACTCCTACCAACTCCGCTTCCCGGACGGTGTCTTCGGCTGCCTCACCCCGCGCGGA 50405
QY 1403 -----GlyPro----- 1404
Db 50406 CCGCGCGCGGACCCCGCTCGCGACACCGCGCGGTCTCGGCGCGCTCGACCCCGCC 50465
QY 1405 ----LeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyAr 1423
Db 50466 TCGTCGCGCTTCGCGCGCGCGGTGTCTTACCAGCGCAACTACGGCGACGGGATCG 50525
QY 1423 qSerIleHisGluIle-----ProArgGluGluLeuArgHisThrPr 1437
Db 50526 G-----CATGCTGCGCAGACGCTTCAGACCCCGCAAGCGCGCGCTCACCG-CC 50578
QY 1437 oGluLeuProLeuAlaProArgProLeuLysGluLysIleThrGlnGly----- 1454
Db 50579 TACTGCGCGCGCGCC---CGGTGACGTGCAATGGAACCCGACGCGGCTCGGAC 50635
QY 1455 -----ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHi 1471
Db 50636 ACCAGGTGCGCGCGCGCTCGCGCTCCACCGCGGACGCGGGGTGTGTTCAC 50695
QY 1471 sAspValArgSerLeuIleGlySerProGly-----ArgThrPheProValHisPr 1489
Db 50696 CACGCGCGTCTTCACAGTCTCGCGCGCGCGCGCTGCGGAGCGCTGTGTGCGCC 50755
QY 1489 oLeuAspValMetAlaAspAlaArgAlaLeuGlu----- 1500
Db 50756 CAGTTCGACGAACCGACCTGCGGAGCCACTCTGCTACGCGGACGCGCGCCATCGAA 50815
QY 1501 -----ArgAlaCysTyrGluGluSerLeuLysSer----- 1510
Db 50816 CCGCGCGTCTATGAGGAACCTGCACCGCTACGCGCGCGCAACTGTTGGCGCGCTGG 50875
QY 1511 -ArgProGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAla----- 1526
Db 50876 CCGGCGCGCAGCTCTCTCTGTCGACAACTCTCTACCGCGCACCGGACGGACCCCTC 50935
QY 1526 ----- 1526
Db 50936 ACCGGGAACCGCGGTGCTGTCGCGCATCGACAGCGCTGGACTGGGACGAGGTGAGC 50995
QY 1527 -----ProValIleVa 1530
Db 50996 GCGTGACCGCGCGCACACCGTGC CGCGACCTTCTGTCACGCGCGCTGTGCGCGT 51055
QY 1530 1-----ProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAs 1545
Db 51056 CCACTCGCCACGCGCGCGGAGGTACCCACGTCGCGCGCTCGCGCTGACCGGG 51115
QY 1545 pHis-----GlyAlaProPheAlaGlyHis----- 1553

Db 51116 ACACCGACACGCGCGCTCACCGAGCGCTCG---GGGGTCAACGCGCGCTCCCG 51172
QY 1554 -----LeuPr 1555
Db 51173 CCTCACCGCCGAACCTCTCCGCGCAGAGAAACCCCGCTTGACCTCCGCGCGGAGCC 51232
QY 1555 oArgGlySerPro---ValThrMetArgGluPro-----ThrProArgLeuGlnG1 1571
Db 51233 CCGAGTCAACCGGTGCGACCTCGCGGAGCCCGTCCGCGCAGCGACCGCTGTGCG 51292
QY 1571 uGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgG1 1591
Db 51293 TGGCGTGTGCGCGCGCGACCGGACCACTCGCGCGCAGCGCACCGCGCGCTTCC 51352
QY 1591 uIleAlaLysSerPro-----HisSerThrValProGluHisHisProHisProlIse 1609
Db 51353 ACCTGTGCGGTCCACGACGACGAGAGGTGTCGCGTTCACGCGCCACACCT----- 51407
QY 1609 rProTyrGluHisLeuLeu-----ArgGlyValSerGlyValAspLe 1623
Db 51408 -CTCTCTGACACACCGTCTCTCTACGCGTGTCTCGCGCGGTCTGCCAGCGTA----- 51461
QY 1623 uTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyLeProLe 1643
Db 51462 ---CGCGCGCGCTTCCG-----CGCGAGCACTACCGCA----- 51494
QY 1643 uAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPro-----Th 1660
Db 51495 -----CGCCACACCTGCGCGACGCGCGCCACGCGCGCTTCCCGTC 51538
QY 1660 rTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeu-- 1679
Db 51539 GGGCGCGCGCTC---CGCGCGCTGTGTGGACCGCGCGCTGCGCGCTGCGCGCC 51595
QY 1680 -----GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHi 1696
Db 51596 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGACCGAAACCCACCGGTGCGCA 51655
QY 1696 sHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProAr 1716
Db 51656 TCCCCGACGCGCGCTGAAAGCCCTGACGCGCTGACCGC---CTGGGGCGCGCGCTCG 51712
QY 1716 gGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerG1 1736
Db 51713 GCGGCAACGCTGCTCGCGCTCATGCGCTGCGC----- 51749
QY 1736 nValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAs 1756
Db 51750 -----CTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51785
QY 1756 pArgLeuAlaTyrLeuProThrAlaProGlnProPhe---SerSerArgHisSerSerSe 1775
Db 51786 -----CGCTTACACCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51814
QY 1775 rProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerG1 1795
Db 51815 ACCACC-TCGACTCGGCGCGCGCGCGCGCGCGCTTACCGACCGCGCTTTCGCGCGC 51873
QY 1795 uArgGluArgAspArgAspArgGluArgAspArgArgAspArgGluArgLysSerIleLe 1815
Db 51874 GACCTCGCGGAGCGC-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51900
QY 1815 uThrSerThrThrThrValGluHisAlaPro---IleTrpArgProGly---ThrGluG1 1833
Db 51901 TTCGGGACGTACGCTGCGCGCGCGCGCGCGCGGTCTTCTGCGCGCGCTGTCGACTACTC 51960
QY 1833 nSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSe 1853
Db 51961 CCCTACGCGCACGCTCGTGGAACTCGGAGGAACTGGCGCGCGCTCACCGCGCGCGCGCA-- 52018
QY 1853 rHisSerHisAlaHisGlnHisSerProHisSerProArgThrGln-----As 1869


```
QY 153 -----ProHisThrAspProGluLeuLeuVal 163
Db 46323 CGAGAGCTACCGCGCGCGTTGCGCCCTCACACCGACTCGG----- 46370
QY 164 ProArgLeuSerLysGluGluLeuLeuGlnAsnMetAspArgValAspArgGluLeu 193
Db 46371 CCCGAGCTGGGACGGCTGG-----GCGCGCGCGCGCGCCCGCAACT 46418
QY 184 ThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGlu 203
Db 46419 CACCCCGCGCGCATGTCGCGCGCTTCTCGAGCTCTCGCCAGTGAGCGACAC 46478
QY 204 AlaAlaLysProGluProGlu-----LysProValSerProProIleGluSer 221
Db 46479 CCGCGCTTCACTCACTCACTCACTTCCACCGCGCGCTCTGCTCCCGCGTGA 46538
QY 222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 241
Db 46539 CGACCTCGTCGGGACTTAC-----CACACGACCTGCTCGGGTGCAGCGGAGGGA 46595
QY 242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
Db 46596 CAC-----CTTCGGGACCGGCGCGCCGCACTCCAGGACGGCATCTGGGAGGACCTGA 46649
QY 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
Db 46650 ACACCGCTGCTAGCGCGCTCGAGTCTCGGATGCTCG----- 46691
QY 282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg 301
Db 46692 -----CCGGAGCGGGGACCCACGACG-----CGT 46718
QY 302 TyrAsp-----GlnLeuMetGluAlaLeuLysLysValGluArgIleGluAsnAsn 319
Db 46719 CCGATCCCGCTGCTTTCACGACACCTTCGCGCGCGCGCGCGCGCGCGCGCGCG 46778
QY 320 ProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db 46779 CCGCGCGCTGCGGTACGCGCGCGCTACGCGCTACGCGACAGACCGCGAGTCTCT-- 46835
QY 340 IleArgLysGlnArgGluGlnGlnArgMetGlnSerArgValGlnArgGlySer 359
Db 46836 -----GCTCGACCATCAGTGAGCGAGCGAGCGCGCGACT 46871
QY 360 GlyLeuSerMetSerAlaAlaArgSerGluHisValSerGluIleIleAspGlyLeu 379
Db 46872 GGTCTGACCTGGACTACGTCGGGACG----- 46901
QY 380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Db 46902 -----CTACCGCGCGG----- 46913
QY 400 TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db 46914 -----GCTGATCGAGCCCATGTTTCG-----GCGCTTCGAGCGCTCCT 46952
QY 420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
Db 46953 CGCTCTCGCGGTACGACGA----- 46976
QY 440 PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu 459
Db 46976 ----- 46976
QY 460 ArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyr 479
Db 46977 -----CGACCGCGCGCACGACGA 46994
QY 480 LysSerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGln 499
Db 46995 CGACCGCGCGCACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47054
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArg----- 514
Db 47055 CGGCCACGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATCA 47114
QY 515 -----SerSerGlnGluLysAspGluLysGluLysGluLysGluLysAlaGlu 530
Db 47115 CGGCCACAGTGCACGACGACGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAC 47173
QY 531 LysGluGluLysPro----- 537
Db 47174 CGGAGTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCT 47233
QY 538 ValGluAsnAspLysGluAspLeuLeuLysGluLysThr----- 550
Db 47234 GCGATGATGAACGACGCGCACCGACCCGTGCGCGCGCACACCTCACCGCCCACTGGC 47293
QY 551 -----AspAspThrSerGlyGluAspAsnAspGluLys 561
Db 47294 CGACGCGCGCGCGCACGACGCGCGCTCTGGCACTGGTGGCACCGGGTCTGACACTGAG 47353
QY 562 Glu-----AlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArg 577
Db 47354 CCACGCGGAACGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 47413
QY 578 -----Arg 578
Db 47414 CATCCCGCGGACCGGCTCGCCCTCGCGTACGCTCGGAGTACGCTGGGAGCAGGCTCGGGCGC 47473
QY 579 LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIle-ThrPr 598
Db 47474 CTTGGCGCGCTCGCGCGCGCGCGCGCGCTCTGCTCGCGCGCGCGCGCGCGCGCGCG 47533
QY 598 oGlnGln-----SerAlaGluLeuAlaSerMetGluLeu 609
Db 47534 CGCCCGCTGCGCACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTCACCCAGTCTGGCT 47593
QY 609 uAsnGluSerSerArgTyrThrGluGluMetGluThrAlaLysLysGlyLeuLeuG 629
Db 47594 CACCCAGCGCATGCACTGGCGCGCGGAA-----CTGCCGCTCTCTCTC 47635
QY 629 uHisGlyArgAsnTrpSerAla-----IleAlaArgMetValGlySerLysThrValse 647
Db 47636 -CGTGACGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47694
QY 647 rGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluLeuG 667
Db 47695 CGACCGACGCGCGCTACCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47743
QY 667 nGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaPr 687
Db 47744 -CACCAACCGCGCGCTGGAGATCGACCGCGCTTCCGCGCTCGGACCGCGCGCGCG 47802
QY 687 oAlaAlaAlaSerGluAlaAlaPheProProValValGluAspGluMetGluAl 707
Db 47803 TGGC-----CCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47846
QY 707 aserGlyValSer-----GlyAsnGluGluMetVa 718
Db 47847 TTCGGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47906
QY 718 lGluGlu-----AlaGluAlaLeuHisAlaSerGlyAsnGluValPr 732
Db 47907 CCGGAGCGCTTGACGAGCGGCTGCGCACCCACCGCGCTCACCTCTGGCACTCGCGCG 47966
QY 732 oArgGlyGluCysSerGlyProAlaThr----- 741
Db 47967 GCCTCTCGCGCTCTCTCTCGACACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48026
QY 742 -----ValAsnAsnSerAspThrGluSerIleProSerPr 754
Db 48027 CTCGCGCTGTGTCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48086
QY 754 oHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProProAlaThrLeuG 774
```

| | | | | | |
|----|-----|--------------------------------------|----------|--------------|----|
| FT | CDS | 12291..15491 | /*tag= f | 12291..15491 | XX |
| FT | | /note= "ORF 25; encodes AAB07561" | | | PA |
| FT | CDS | 15488..21013 | /*tag= g | | XX |
| FT | | /note= "ORF 24; encodes AAB07562" | | | XX |
| FT | CDS | 21010..24666 | /*tag= h | | DR |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | DR |
| FT | | /note= "ORF 23; encodes AAB07563" | | | DR |
| FT | CDS | 24663..32690 | /*tag= i | | XX |
| FT | | /note= "ORF 22; encodes AAB07564" | | | XX |
| FT | CDS | 32893..34830 | /*tag= j | | PT |
| FT | | /note= "ORF 21; encodes AAB07565" | | | PT |
| FT | CDS | 34827..35804 | /*tag= k | | XX |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | XX |
| FT | | /note= "ORF 20; encodes AAB07566" | | | CC |
| FT | CDS | 35818..37302 | /*tag= l | | CC |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | CC |
| FT | | /note= "ORF 19; encodes AAB07567" | | | CC |
| FT | CDS | 37299..39215 | /*tag= m | | CC |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | CC |
| FT | | /note= "ORF 18; encodes AAB07568" | | | CC |
| FT | CDS | 39301..47181 | /*tag= n | | XX |
| FT | | /note= "ORF 17; encodes AAB07569" | | | SQ |
| FT | CDS | 47178..49985 | /*tag= o | | |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | |
| FT | | /note= "ORF 16; encodes AAB07570" | | | |
| FT | CDS | 49982..51001 | /*tag= p | | |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | |
| FT | | /note= "ORF 15; encodes AAB07571" | | | |
| FT | CDS | 50998..52386 | /*tag= q | | |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | |
| FT | | /note= "ORF 14; encodes AAB07572" | | | |
| FT | CDS | 52383..52946 | /*tag= r | | |
| FT | | /note= "ORF 13; encodes AAB07573" | | | |
| FT | CDS | 53018..54190 | /*tag= s | | |
| FT | | /note= "ORF 12; encodes AAB07574" | | | |
| FT | CDS | 54187..55824 | /*tag= t | | |
| FT | | /note= "ORF 11; encodes AAB07575" | | | |
| FT | CDS | 55821..56093 | /*tag= u | | |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | |
| FT | | /note= "ORF 10; encodes AAB07576" | | | |
| FT | CDS | 56090..57586 | /*tag= v | | |
| FT | | /transl_except= (pos: 1..3, aa: Met) | | | |
| FT | | /note= "ORF 9; encodes AAB07577" | | | |
| FT | CDS | 57583..58857 | /*tag= w | | |
| FT | | /note= "ORF 8; encodes AAB07578" | | | |
| XX | | | | | XX |
| PN | | WO200040704-A1. | | | |
| XX | | | | | XX |
| PD | | 13-JUL-2000. | | | XX |
| XX | | | | | XX |
| PP | | 06-JAN-2000; 2000WO-US000445. | | | PP |
| XX | | | | | XX |
| PR | | 06-JAN-1999; 99US-0115435P. | | | PR |
| PR | | 05-FEB-1999; 99US-0118848P. | | | PR |
| PR | | 05-JAN-2000; 2000US-00477962. | | | PR |

XX (REGC) UNIV CALIFORNIA.
PA Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
PI
XX
XX
DR WPI; 2000-465974/40.
DR P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
DR AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR AAB07576, AAB07577, AAB07578.
XX
XX New bleomycin gene cluster components useful for peptide and/or
PT polyketide metabolites, especially bleomycin, production and for
PT chemically modifying biological molecules.
XX
XX Claim 8; Page 97-136; 162pp; English.
XX
CC The present sequence represents the BLM (Bleomycin) gene cluster,
CC containing open reading frames (ORFs) 8-30. The proteins encoded by the
CC gene cluster are useful for producing peptides and/or polyketide
CC metabolites, especially bleomycin or bleomycin analogues. They are also
CC useful for chemically modifying biological molecules to produce branched
CC methyl groups, and for coupling amino acids and fatty acids. They may be
CC reacted with an apo-carrier protein and coenzyme A to produce a holo-
CC carrier protein. The BLM gene cluster or catalytic domains can be used
CC individually or collectively to produce thiazolidine, thiazoline,
CC bithiazoline and bithiazoline-containing microbial metabolites. The BLM
CC gene cluster may also be used to produce sugars
XX
SQ Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,29e-11 Length: 58857
Score: 666.00 Matches: 681
Percent Similarity: 30.64% Conservative: 275
Best Local Similarity: 21.83% Mismatches: 1098
Query Match: 5.04% Indels: 1078
DB: 3 Gaps: 156

US-09-522-753-5 (1-2517) x AAA58471 (1-58857)
QY 3 GlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProHis 22
Db 46008 GGTGCGGACGCGCTCCGCTGGCGCTGTGACGTACGACCA-----CGCTGGA 46061
QY 23 SerLeuSerTyrProVal-----GlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 46062 CGACGTCCGACCCCGCTGCACCTGAGCTGGACCTGCTCAT----- 46103
QY 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleLeuGln 60
Db 46104 -----CGCGACGCCACAGCGTCCACGTACTACCGCGGACCTCTCCTTCTACGC 46157
QY 61 ProGlnArgArgProSer-----LeuLeuSerGluPheGlnProGlyAsn 76
Db 46158 CGACCCCGACGCGCCCTGCGGCTCTCTTCCGCGACTACGCTCTTCTACGC 46217
QY 77 GluArgSerGlnGluLeuHisLeuA-gProGluSerHisSerTyrLeuProGluLeuGly 96
Db 46218 CCGCGC-----CCACCGCGAGGG 46235
QY 97 LysSerGluMetGluPheIleGluSerLysArgProArgLeuLeuLeuProAspPro 116
Db 46236 CGA-----GCCGCGCCCGCGC-----CCTCGACCA 46262
QY 117 LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThr 136
Db 46263 CTGGCGGCGCGGCT-----GGCGACCTCCCGG----- 46292
QY 137 LysAspArgSerLeuThrGlyLysLeuGluProValSerProSer----- 152
Db 46293 -----CCGCGCGGCTGCGCTGCGGCGCGCC 46322

[illegible][illegible]

```
Qy 1311 TyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeu 1330
Db 98250 GCCGGCCCGCCGACGGTGTATACCAATGCGCGCTT---GCCGGCCCGCCGCAAGCC 98306
Qy 1331 MetGlyArgAlaIleProProGluArgHisSerProHisLeuLysGluGlnHisHis 1350
Db 98307 GCCGGCCCGCCGA-GTCCGCGGTGCGCACCGACCCACCGTGTGCTGACTGCGGA-- 98362
Qy 1351 IleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyr 1370
Db 98363 -----CCGTCCCGGAAGTGGCGCCCGCCGACCCAC 98395
Qy 1371 LeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProSer 1390
Db 98396 CGTCCCGCGCATGTCCACCGCCCGCGGACCG-CCATGCGCGCCGACACCGCCGCGCC 98454
Qy 1391 ArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAla 1410
Db 98455 CCACCGGTGCGCGATACTGTGCGATA--CCGTGGCGGCCCTG----- 98496
Qy 1411 HisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArg 1430
Db 98497 -----GCCCGCGCC 98505
Qy 1431 GluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly-- 1449
Db 98506 CCACCGGTAGGCCCAACCGCGTTCGCGCGGCCCG--CCGTGACC-GCCGGACCC 98561
Qy 1450 SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLys 1469
Db 98562 GCCGTACCGCGACACCGACCC-----ACCGCGCCAC----- 98594
Qy 1470 LysHisAspValArgSerLeuIleGlySerProGlyArgThr-PheProProValHisPr 1489
Db 98595 -----GGCACCCCGCGCACCCACCGCGCCGCGCGCTGCC 98630
Qy 1489 oLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLy 1509
Db 98631 C----- 98631
Qy 1509 sSerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaProValIl 1529
Db 98632 -GCGTCGCCCTGACCCCGTTCCTCCCGTGGCCGAGCGGAGGCGATCT----- 98682
Qy 1529 eValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGlnAspHisGlyAlaPr 1549
Db 98683 -----GAACGAGCGCGCCGAATTGGCCCGTGGCGCGCGCGCATCGCCACC 98735
Qy 1549 oPheAlaGlyHisLeuProArgGlySerProValThrMetArg-----GluProTh 1566
Db 98736 A-----TTGCGCGCGCGCCACCATCGCGACCCGCGCGCATTCGCGCGTC 98783
Qy 1566 rProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuTh 1586
Db 98784 GCCTCGGTGCC----- 98798
Qy 1586 rSerThrProArgGluLeuAlaLysSerProHisSerThrValProGluHisHisProHi 1606
Db 98799 GGCGCGCGCGCGACGTGGCTTCGCGACCGTTCGCCACCGTTACACCGTTCGCGCGCT 98858
Qy 1606 sProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgse 1626
Db 98859 GCCG--GGCCCGTGG-----TCGCTGGGTGCCACCGCTG---GCCGC 98897
Qy 1626 rHisIleProLeuAlaPheAspProThrSerIlePro-ArgGlyIleProLeuAspAlaA 1646
Db 98898 GCGGGTGGCGTTCGCGTGTTCGCGCGCCCAAG---CCGCGAGGCGCTCCGCGCGCGGGTC 98954
Qy 1646 laAlaAlaTyrTyrProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrP 1666
Db 98955 ACCGCGGTACCGCGCGTCCCGCGCATCA-CTGCCATGACCGCGCTCACCGCGGTGCCG 99013
```

```
Qy 1666 ro-----ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnA 1682
Db 99014 CGGTCGCGAGCCGCGC-GTCACCGCAACCGCGCGACACCGTGGCT----- 99062
Qy 1682 rgGlnThrIlelleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrA 1702
Db 99062 ----- 99062
Qy 1702 laMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaL 1722
Db 99063 -----GCCGACCTACCGCGGGTGGCGCCCGCGCGCGCGCATC----- 99104
Qy 1722 euAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProV 1742
Db 99105 -----ACCGCGCGGCGCC----- 99116
Qy 1742 alLeuVal-ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeu 1761
Db 99117 --GCCGTACCGCAACCGCGCGACACCGCGCTCGCG----- 99153
Qy 1762 ProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGly 1781
Db 99154 CTTTCCCGCGCGCGCTCCACCGGCTCAGCGCTGTGCGCGCGGAGGCGCC----- 99207
Qy 1782 ProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAsp 1801
Db 99208 -----TTGCGCGCGCGCGC-ACCCTAGTCTCGCGC----- 99236
Qy 1802 ArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrVal 1821
Db 99237 -----ACCACCGCGACCGCGCGCGCGCTTCCACCGGGGTCCACCGTCGCC 99284
Qy 1822 GluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSer 1841
Db 99285 CGACACCGACCC-----ACCGCGCGCACCGGCTCGGCC 99317
Qy 1842 GlyGlyGlyGlyGlySerSerSerArgProAlaSerHis----- 1854
Db 99318 -----GGCACCGCGCGCGC-CGCGCTGCCCGCATCGCTCGACCGCGTTCGCGC 99370
Qy 1855 -----SerHisAlaHisGlnHisSerProIleSerProArg----- 1866
Db 99371 CACTACTAACCGCGAACTCCCGCGCACCGCTACCGCGCGCACCGCGCGCGCGCAC 99430
Qy 1867 ThrGlnAspAlaLeuGlnGlnArg---ProSerValLeuHisAsnThrGlyMetLysGly 1885
Db 99431 CGCAACCGCGCTTGACCGCGCTTGGCTTCGCTTCCTTCGCGCGGTGGCCGAGAGG 99490
Qy 1886 IleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerSerSer 1905
Db 99491 TGCGG-TCGGCG-----CCGTCTCCTACCTACGTTC-----GCG 99522
Qy 1906 ProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeu 1925
Db 99523 CCATTGCGACCGCGCGCGCTTACACCGTTCGCGCGCGCGCGCGCGCGCACCAAC 99582
Qy 1926 AspGlyValTyrProThrLeuMetGluProValLeuLeuPro-----LysGluAla 1942
Db 99583 CGCGCGTTACACCGCGCACCGCGCTTCGCGCGCGCACCGCGCGCGTTCGCGC 99642
Qy 1943 ProArgValAlaArgProGlu-----ArgProArgAlaAspThr 1955
Db 99643 CCGTTACCGCGCTTCGCGCGTTCGCCACCATCTGACGCTAGCGCGCGTTCGCGCGGTC 99702
Qy 1956 GlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPro 1975
Db 99703 ACACCTCTCGTGGCGCGCGCGCGCGCG-----CCACCTCTCCACCGC 99747
Qy 1976 SerLysGlySerGluProArgProLeuValProValSerGlyHisAlaThrIleAla 1995
Db 99748 GCGCGTGGGTACACCATCGCGCGCGCGCGCGCG-----TCACCATCTCTGAAAGCC 99801
Qy 1996 ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProPro 2015
```


| | | | | | | | |
|----|-------|--|-------|----|-------|--|-------|
| Db | 96394 | T-----GCCCGTCTCGCTCGCTGCAATGTGAGTCCCGCAGTGC----- | 96433 | Db | 97325 | TGCCGACCTCGCGCGGGTGGCGAGCCCGCGC----- | 97363 |
| Qy | 654 | eAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetG1 | 674 | Qy | 985 | roArgGluAspAlaAlaProThrLysProAlaProProAlaProProGlnAsn- | 1004 |
| Db | 96434 | -----TGGCAGCATCGCCCGAAGACC | 96456 | Db | 97364 | CGTTTCGCGCGCGCGCCCGCCCGCGTAGCCACCGTGGCGCGCGCGCGTGC | 97423 |
| Qy | 674 | uLysGluArgAsnAlaArgLysLysLysAlaProAlaAlaAspSerGluGluA1 | 694 | Qy | 1005 | -----LeuGlnProGluSerAspAlap | 1012 |
| Db | 96457 | GACACGACGCATCGTGGGGGTG-CGTCGAAGTCGCGGTGACCGATCGCGGAGTC | 96515 | Db | 97424 | GGCCCGCTGTTTGTAAATGTCAAAGCCGTCAACGCCGTATACCGCGACCCCAAGCC | 97483 |
| Qy | 694 | a-----AlaPheProValValGluAspGluGluMetGluAspSerGlyVa | 710 | Qy | 1012 | roGlnGlnProGlySerSerProArgGlyLysSerArg-----SerProAlaProp | 1029 |
| Db | 96516 | ACCGGCCGATGGGGTTTCGAAGGCTAGGGAATGATGACGATGGGGCGCCCTCGGC | 96575 | Db | 97484 | CGCTAGGCTCGCGCCCGCGCTACACCGTGGCGGTCTGAGTTCGCGCGCGCCAC | 97543 |
| Qy | 710 | 1-----SerGlyAsnGluGluGluMetValGluGluAlaG1 | 722 | Qy | 1029 | roAla-----AspLysGluAlaPheA | 1036 |
| Db | 96576 | CGCTTCGCGTAAACCCCAACCATGCGGAACAGACCTAGCTGCGCGCGCGCGCAG | 96635 | Db | 97544 | CGCGCCCGCGTGCCTCCCGCCCGCCCGCGTGGCGGTGCGTTCGCTCCCTGCGC | 97603 |
| Qy | 722 | uAlaLeuHisAlaSerGly---AsnGluValProArg-----GlyG1 | 735 | Qy | 1036 | laAlaGluAlaGlnLysLeuProGlyAspPro-----ProCysTyrThrSerG | 1052 |
| Db | 96636 | AGCGAGCATCGCGTGGGCCCAACGACGAGGATCGGATCTCTGACCATGGTGG | 96695 | Db | 97604 | GCTCGAGTGGCGGACCAACCGCGCCACCGCGCGCGCGTTCCTCCCTGCGC | 97659 |
| Qy | 735 | uCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHi | 755 | Qy | 1052 | lyLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaAspP | 1072 |
| Db | 96696 | GTCTTCTGGTACGTGACGTTA-----ACGAGATGAGGGCGCGCTTCGCGC-- | 96744 | Db | 97660 | -----CCGTTCCGCGCGCGCGTTCATCCCGCGTCCCGCGCGCGCGCGC | 97707 |
| Qy | 755 | sThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyA1 | 775 | Qy | 1072 | roSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrA | 1092 |
| Db | 96745 | -----CCTTCACCGCGC----- | 96756 | Db | 97708 | -----TGTCCACCAACCCCGCGCGCGCGCGCGCGCTG----- | 97740 |
| Qy | 775 | aAspGlyProProGlyProProThrProProArgArgThrSerArgAlaProIleG1 | 795 | Qy | 1092 | laArgProValLeuProArgProProThrIleSerAsnProProPro-LeuIleSerSer | 1111 |
| Db | 96757 | -GAACCGCGGAGCGAGGTGCGCCCTCCGTTGCGGAGCCACCGACTCGCGCAGCA | 96815 | Db | 97741 | -----CGCGCCCGCGCAGCCACCTTGG-----CGCGCGCTCCCGCGTCC | 97783 |
| Qy | 795 | uProThrProAlaSerGluAlaThrGlyAlaProThrPro-----ProProAlaProPr | 813 | Qy | 1112 | AlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVal | 1131 |
| Db | 96816 | GCGAGCCCGCGAGGTCAAGCCACCGGACCGCGCTCCCGCTCACCTCCGCGCGCGC | 96875 | Db | 97784 | CGCGTGGCCAGAAATGCTCCCGCGCGC----- | 97813 |
| Qy | 813 | oSerProSerAlaProProValValProLysGluGluLysGluGluGluThrAlaA1 | 833 | Qy | 1132 | GlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGly | 1151 |
| Db | 96876 | GTCCCGCGCTACCGCGCGGATGCTGCG-AGCGAGGGGGCGCGAAGCCCGCGAGC | 96934 | Db | 97814 | -----CACAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC | 97866 |
| Qy | 833 | aAlaProPro-----ValGluGluGluGluGluGluGluLysProProAlaA1 | 849 | Qy | 1152 | LeuPro---LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlu | 1170 |
| Db | 96935 | CGCGGTCCCGCTCGCTCCGTCACACCGGGGGCGCCCGCTCTCGCGCGCGCCAC | 96994 | Db | 97867 | CTTCGAGGCGACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC | 97911 |
| Qy | 849 | uGluLeuAlaValAspThrGlyLysAlaGluProValLysSerGluCysThrGluG1 | 869 | Qy | 1171 | GlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGln | 1190 |
| Db | 96995 | CCAAGCGCGCTT---GCCGCGTGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGC | 97051 | Db | 97912 | CCCGCTCACCGATGTCGCGCGCGTCAACCGCGC-----TTGCTCCAGCGCCCA | 97959 |
| Qy | 869 | uAlaGluGluGluProAlaLysGlyLysAspAlaGluAlaAlaAlaThrAlaGluG1 | 889 | Qy | 1191 | GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys | 1210 |
| Db | 97052 | TGCGGTGCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT | 97111 | Db | 97960 | CGCGCCCC-GCCATCACCGCGAGAGCTCGGAGCGGTGCGCGCGCGCGCGCGCGCGC | 98018 |
| Qy | 889 | YAlaLeuLysAlaGluLysGluGlyGlySerGlyArgAlaThr-----Th | 905 | Qy | 1211 | GlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThr | 1230 |
| Db | 97112 | CGCC-----GAAGCGGCTTGGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGC | 97156 | Db | 98019 | GGCTCGCGTCCCGCAATAGCCCGGTGGCGCGTCAACCGCGCGCGCGCGCGCGCGC | 98078 |
| Qy | 905 | rAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAlaAspG1 | 925 | Qy | 1231 | HisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAsp | 1250 |
| Db | 97157 | CGCTTTCGC | 97216 | Db | 98079 | GCGCGCGCGCGCGC-----CTTGACCGTTCGCGC----- | 98108 |
| Qy | 925 | uValAspGluAlaGluGlyLysAspLysAsnArgLeuLeuSer-ProArgProSerLeuL | 945 | Qy | 1251 | SerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyr | 1270 |
| Db | 97217 | GCTC---TCCGTCAAAATCCGCGAGATAGTAAAGCCGTCAACCGCGCGCGCGCGCGC | 97273 | Db | 98109 | GCGAGCGCGCGTCCCGCGCGTCTAGTGGCGCGTTCGCTGGCGTCCAGCGCGT--- | 98165 |
| Qy | 945 | euThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysG | 965 | Qy | 1271 | GluGlyLysLysGlyHisValLeuSerTyrGluGlyMetSerValThrGlnCysSer | 1290 |
| Db | 97274 | CAGGTCCCGC | 97324 | Db | 98166 | ---GGCGCGGAGGTGCGGTGGCGTGTCAACCGCGCGCGCGCGCGCGCGCGCGCGC | 98222 |
| Qy | 965 | lnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProp | 985 | Qy | 1291 | LysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThr | 1310 |
| Db | | -----CGAT | | Db | 98223 | -----GGGTACACCGCGTTCACCGCGCGCTCC | 98249 |


```
Qy 1684 -----ThrIleAsnAspTyrIleThrSerGlnMethHisAsnT 1699
Db |||:||||:
51642 CTGACGACCAACATACCTGGTGTAGCCAGCATCCAGCGGTGTGAGACTCACCCGTCAA 51583
Qy .1699 hrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerS 1719
Db |||:||||:
51582 CGTCGACACGACCGCGAGGTGC-----AACGCGGTAGTCCACGC----- 51542
Qy 1719 erLeuAlaLeuAsnTyr-----AlaAlaGlyProArgGlyIleIleAspL 1734
Db |||:||||:
51541 -----TGCCTGCGATATCTGTGAATCTCTGAGCATCGGTGTCCATCGATACGAATGAAA 51487
Qy 1734 eu-----SerGlnV 1737
Db |||:||||:
51486 CGCATGTGAACCCGGAACCAACCCGCCACCCCTTGGCCATCCAGCCGCGCGCAAT 51427
Qy 1737 alProHisLeuProVal-----LeuValProProThr----- 1747
Db |||:||||:
51426 GTCTCAAGCACCTCCCGATCGCGGAGAGACCAACCCGACTCAGAAGCGTTGACCGCGGC 51367
Qy 1747 ----- 1747
Db |||:||||:
51366 GATCCCAACCCCTTCAACCCGAGCGGCAACCAACGCGCGCAGTGTCTCTCGCCAGCAGC 51307
Qy 1748 -----ProGlyThrProAlaThrAla----- 1754
Db |||:||||:
51306 CACCGCCAAATGCGACCGCCATCCGCAACCGCTGTATCAACCGCGGACGACCCGCCAC 51247
Qy 1755 -----MetAspArgLeuAlaTyrL 1761
Db |||:||||:
51246 CAACGAGCGCATCCGCGAACGACACACACCCCGCGCAACCGCGCAAACTCAC 51187
Qy 1761 euProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSer----- 1778
Db |||:||||:
51186 CACCGAATGCCATCCATCACACGAGCGCAACACCCACGAGCCAGCGCTTCAGCAA 51127
Qy 1779 -----ProGlyGlyProThrHisLeuThrLysProThrThrSerSerSerGluArgG 1797
Db |||:||||:
51126 ACCAACTGCMACCGGAACAAACCGGACTGGCGCCCAACACGTCCGATCCAGCAAGGACCC 51067
Qy 1797 luArgAspArg----- 1800
Db |||:||||:
51066 GTCGGAACCGGAACACCATCCGAAACCCACCTGCTGCCCCAGATGCGCATCCAGTCC 51007
Qy 1801 -----AspArgGluArg-AspArgAspArgGluArgGluLysSerIle 1814
Db |||:||||:
51006 GGCAAAAGCTTCGTCAAACCGCTGAGCAACACCGGAACCGCGAGTAGAGCCCTCTTCC 50947
Qy 1815 -----LeuThrSerThrThrValGluHisAlaProIleTyrArgProGly 1830
Db |||:||||:
50946 CATCCCGACCACTGACC-----ACCTGACCGGAACACAAACCC-----AACCCCGCC 50896
Qy 1831 ThrGluInSerSer-----GlySer-SerGlySerSe 1841
Db |||:||||:
50895 AGAGATCAGGATCCCGACACACCCAGAGCGTTCGCGACCGCGACGCGCTTCAA 50836
Qy 1841 rGlyGlyGlyGlySerSerArgProAla-----SerHisSerHisAl 1857
Db |||:||||:
50835 CCCAGACAACAGCGCTTCACGCTCGCTCCAGCAGCAGCCGCGCTCCTCCAAACGCGGC 50776
Qy 1857 aHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSer-- 1876
Db |||:||||:
50775 CCGGGAAGACCAACGAAACCGCGGCATCAGGGATGAATATCGGGACGATTATCAA 50716
Qy 1877 -----ValLeuHisAsnThr-- 1881
Db |||:||||:
50715 ATAGACATCAATGCGCTGCGCTGAGCGCTGAGCGCTTCCGGTGTCTCTGCGCGACAGAT 50656
Qy 1882 -----GlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu----- 1898
Db |||:||||:
50655 CAACGGGACCGCTGGGATGTCCACCGTAGATTCCGATTCGGGACCGCTGTGCTCTCGCAAT 50596
```

```
Qy 1899 -----ArgSerThrSerThrSerSerProValArgPr 1909
Db |||:||||:
50595 CTGCTCTGGTGATTCGAGGCTTGTTCGAGATGACATGTCGCTGTGTGCGGTGATGCC 50536
Qy 1909 o-AlaAlaThrPheProPro-----AlaThrHisCys----- 1919
Db |||:||||:
50535 GAACGACGACATCCCGCCGCGAGGATGCGCTGTGCGGCGAGCGGTGTCTCCGT 50476
Qy 1920 -----ProLeuGlyGlyThrLeuAspGlyValTyrP 1930
Db |||:||||:
50475 AAGAGCTGACCTCCCGCCGCGACCAATCCACTT-----C 50440
Qy 1930 roThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluA 1950
Db |||:||||:
50439 CGGCTAGGTTCATCCATGCAACGTCGCGGCGAGCTGCTGTCGCGCATCGCCATCAC 50380
Qy 1950 rgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuG 1970
Db |||:||||:
50379 CATCTTGATCACACCAACACCCCGCTG-----CGCCTGCGTATGACCAATATT 50329
Qy 1970 luProAlaSerSerProSerLysGlySerGluProArgProLeu----- 1984
Db |||:||||:
50328 CGACTTCACGACCCCAACCAACAGCGCCGATCCCTATCCGCTCTGCGCGTAAGTGGC 50269
Qy 1985 -----ValPro-----ProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnL 2002
Db |||:||||:
50268 AATCAGTGCCTGCGCTCAATCGATCACCAAGCTGTGTCCTCCATGCGCTCCAC 50209
Qy 2002 euAlaProHisHisAlaSerProAspProProAlaPro-----ProAlaS 2017
Db |||:||||:
50208 AGCGTCCACATCAGACACCGACACCGCACTCGCAACGCTGCGTGTATCACCCGCTG 50149
Qy 2017 erAlaSerAsp-----ProHisArgGluLysThrGlnSerL 2029
Db |||:||||:
50148 CTGCGACGACCATTCGCGCGCGCTCAACCATTCGACGACCATCTTGATTCACCGCGCT 50089
Qy 2029 ysProPheSerIleGlnGluLeuArgSerLeuGlyTyr----- 2043
Db |||:||||:
50088 ACCCGCACGACCGCCAGAACCGGTGCGCTGTCGCGCGGCATCCGACAGCGCTCAA 50029
Qy 2044 -----HisGlySerSerTyrSerProGlyValGluProValSerProValSerSer- 2061
Db |||:||||:
50028 CAACACGAGCGCGCTTTCGCGCCCGCGTGTGCTATCCGACGACGCGCAACAGCTT 49969
Qy 2061 ----- 2061
Db |||:||||:
49968 ACACGCGCTGCGCGCCCAACCCCTCTGCGGGAAACCTCCACGAACATCCCGCGCT 49909
Qy 2062 --ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSerH 2081
Db |||:||||:
49908 GGCAATCACCGTCACAC-----CACCGCAAGAGCCAGATCACACTC 49867
Qy 2081 isLeuGluGlyGluLeuArgProLysGln----- 2090
Db |||:||||:
49866 ACCAGACGCGAGTGTCTGA-----CCGCGCAGGTGCATGACACCAACGACGACGACGCGC 49808
Qy 2091 -----ProGlyProValLysLeuGlyGlyGluAlaAlaHis-----L 2103
Db |||:||||:
49807 TATCCACCGTCACGCGAGGACCTCGAAACCAACGAATAGCGACCCGCGCGCGAAAGA 49748
Qy 2103 euProHisLeuArgProLeuProGlu-----SerGlnProSerSerSerProLeuLeuG 2121
Db |||:||||:
49747 CGCGCGCGCGCTGCCATTACCTAGATAACCTCGAAACCTTCGCGCGCTGTGTGACGA 49688
Qy 2121 lnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIle---- 2139
Db |||:||||:
49687 ACCGCGCGCGCTA-GTCTGTGTGTCATCAGCC-----CGCGAACACCGCGCTC 49641
Qy 2140 -----SerGluV 2142
Db |||:||||:
49640 CGACTCCCTCGCAACGAAGGATCAATCCCGCGCGCTCCCAAGGTCTCCCAAGCGATC 49581
Qy 2142 alIleThrGlnAspTyrThrArgHisProGlnGlnLeuSer-AlaProLeuProAla 2161
```

```
Db 53578 CCAACTCGGTGATTCGGGAGCATCCAGC-----CAC 53546
Qy 1127 GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly 1146
Db 53545 GACGGCCCGTCA-----ACNCACTGACAGAACACCGCGTTCGGCCACCAAGGT 53498
Qy 1147 ProValThrMetGlyLeuPro-----LeuProMetAspPro-----LysLysLeuAla 1162
Db 53497 -----GGCGTGCACACGACCCACCAACAGACCGGTGCCACCGAAATCAGCA 53450
Qy 1163 ProPheSerGlyValLysGlnGluLeuSerProArgGlyGlnAlaGlyProProGlu 1182
Db 53449 CCG-----TCCGGT-----CGGACCGCGCCACCAACCGCGGCCTCGACCGCGCCGGAA 53399
Qy 1183 SerLeuGlyValProThrAlaGlnGluAla-----SerValLeuArgGlyThr 1198
Db 53398 -----CCGACGCCACACCGCGGCACTGTCCACCAACCGCTTACGACCGCCCACT 53351
Qy 1199 AlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPro 1218
Db 53350 GCGGTTGCATCGACCCACCGCACTCGGCAACGCGGCCACGACTCGGGGTACCGTCCA 53291
Qy 1219 SerAspSerAlaIleThrTyrArgGlySerIleThr----- 1230
Db 53290 CATCGACCAACAAACCGCGCGGTTCCTCCGACTCGCTGAAACGCAACACCCCCACA 53231
Qy 1231 ----- 1232
Db 53230 ACGCGCATCGCCAGTTTGATGCGCGCTCACCGCAGCGGTCTGTCACCGCCACAG 53171
Qy 1233 ThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerPro 1252
Db 53170 TCACCAACACGACGACCGCGCCCAACCGCTCTCTCG-----CCA 53129
Qy 1253 SerArgLeuAspArgGlyArgGluAspSerLeuPro----- 1264
Db 53128 ACCACGACTGACGACGCGCGCAGCAT-----CTCCCGGTCTCCGCGAGTCCGCCACGC 53075
Qy 1265 -----LysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyG1 1283
Db 53074 CCGGTCCGATCGCGCTGTTACCGCGCGCAATAAACACGACATCAGCAACACG----- 53020
Qy 1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProHisG1 1303
Db 53019 -----CGCACCGGAATCAAGTTCCCGCGA 52994
Qy 1303 uThrAlaAlaProLysArgThrTyrAsp-----MetMetGluGlyArgValGlyAr 1320
Db 52993 GCTCAGCCACACCGCACACGATCGACGGAACCTGTCGCGAGCCCAATGAGACCGGCCA 52934
Qy 1320 gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHi 1340
Db 52933 TGCGCGCAGCAGATACCAAGAACCGCCAGGACCCGATACCGTCGACCGGTGCCA 52874
Qy 1340 sSer---ProHisLeuLysGlnGlnHisIleArgLysIleThrGlnGlyI1 1359
Db 52873 CCGGCTCCCAAGTCACCTCGAACAGCG-----GTCACCGCAACAGT 52832
Qy 1359 e-----ProArgSerTyrValGluAlaG 1367
Db 52831 CCGGATTCGCCAACCGCACCGACACCGCGCGCAACCGGAGAGATCATCGAGCGCA 52772
Qy 1367 lnglu-----AspTyrL 1371
Db 52771 GCACACCGCCACCTCATTCATCAGCCACCGACCGGATCGGTTCGGGTAGATCCA 52712
Qy 1371 euArg-----ArgGluAlaLysLeuLeuLysArgGluG 1382
Db 52711 CCGGTGCGCTGCTGATCGTCCACCTGACCGGAAAGTTCTACCCCGGTAAGAGAACG 52652
Qy 1382 lYThrProProProPro-----ProProSerArgAspLeuThrGluAlaTyrL 1398
```


| | | | |
|------------|---|--|------|
| QY | 2453 | rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLeuMetArgLeuGlnA | 247 |
| Db | 1559 | CACCCGCCAGCAGGGGTGGCCGCCCC | 1534 |
| QY | 2473 | laGlyValMetAlaSerProProProPro-GlyLeuPro---AlaGlySerGlyProLeu | 2491 |
| Db | 1533 | --GGGCGAGCGCAGCTCGGCGGCGAGCGGTCTGCACCTTCAGCCGCGAGCTGGTCCAGCG | 1476 |
| QY | 2492 | AlaGlyProHisAlaTrp | 2498 |
| Db | 1475 | GGATGCCCGAGAGAGGGGTGG | 1455 |
| RESULT 34 | | | |
| ABV75557/c | | | |
| ID | ABV75557 | standard; DNA; 75236 BP. | |
| XX | AC | ABV75557; | |
| XX | AC | | |
| DT | 22-JAN-2003 | (first entry) | |
| XX | | Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 1. | |
| XX | | Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide; | |
| KW | | metabolite; spinosyn; gene; ds. | |
| KW | | | |
| XX | | Saccharopolyspora sp. | |
| OS | | | |
| XX | Key | Location/Qualifiers | |
| FH | CDS | 1..13035 | |
| FT | | /*tag= a | |
| FT | | /product= "busA" | |
| FT | | 13059..19508 | |
| FT | | /*tag= b | |
| FT | | /product= "busB" | |
| FT | | /note= "No start codon given" | |
| FT | | 19553..29056 | |
| FT | | /*tag= c | |
| FT | | /product= "busC" | |
| FT | | 29092..43893 | |
| FT | | /*tag= d | |
| FT | | /product= "busD" | |
| FT | | 43945..60639 | |
| FT | | /*tag= e | |
| FT | | /product= "busE" | |
| FT | | 62090..63940 | |
| FT | | /*tag= f | |
| FT | | /product= "ORF RI" | |
| FT | | /transl_except= (pos:62879..62881,aa:Xaa) | |
| FT | | /transl_except= (pos:62939..62941,aa:Xaa) | |
| FT | | /note= "Xaa=Unknown. No start codon given" | |
| FT | | complement(65226..66602) | |
| FT | | /*tag= g | |
| FT | | /product= "ORF RII" | |
| FT | | /note= "No start codon given" | |
| FT | | complement(68759..69676) | |
| FT | | /*tag= h | |
| FT | | /product= "ORF RIII" | |
| FT | | /note= "No start codon given" | |
| XX | | | |
| PN | WO200279477-A2. | | |
| XX | | | |
| PD | 10-OCT-2002. | | |
| XX | | | |
| PF | 28-MAR-2002; 2002WO-US009969. | | |
| XX | | | |
| PR | 30-MAR-2001; 2001US-0280175P. | | |
| XX | | | |
| PA | (DOWC) DOW AGROSCIENCES LLC. | | |
| XX | | | |
| PI | Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C; | | |
| PI | Mitchell JC; | | |
| XX | | | |
| DR | WPI; 2003-058434/05. | | |

Db 4337 CGTCTGGTTGACCGCACTCCCGCGCACCGCCAGCACCC----- 4296
QY 1547 GlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArg----- 1563
Db 4295 GGTGTCCCGCAGCCGGGCGT---CGGAGAGCGCTCCAGTGCAGTACGCGCACCCCT 4239
QY 1564 -----GluProThrProArgLeuGlnGluGlySerLeuSerSerHisAlaSerGln 1581
Db 4238 CGGACAGGCGGTACCGTCCGATCGCGGAGAACGATGACGCGCGCTCGGGGCGGA 4179
QY 1582 AsArgGlyLeuThrSerThrProArg-----GluLeuAlaGlySerProHisSerThr 1599
Db 4178 GCCCGCGCTCCCGAGAACTCCACGAACATGCGGCGCCCGACATCACCGTGGCGCCG 4119
QY 1600 ValProGluHisProHisProHisPro-----TCGCCCCGGCGAGCGCTCCAGCGGCAGGTGC 1610
Db 4118 CGGCCAGAGCC-AGCCGCGAC-----TCGCCCCGGCGAGCGCTCCAGCGGCAGGTGC 4066
QY 1611 -----TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArg 1625
Db 4065 AGCGCCACAGGAGCAGGAGCAGCGGTGTGACGCGTGAGC----- 4024
QY 1626 SerHisLeuProLeuAlaPheAspProThrSerIleProArg-----GlyIleProLeu 1643
Db 4023 -----GCCGTCCGCGC-----GTGCCGAGGTGTAGCGATCCGCGCG 3985
QY 1644 AspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHis 1663
Db 3984 GAGCCACGCTCGG-----GTGGTCCGCGTGAGCAGTACCCCTCG 3943
QY 1664 LeuTyrPro-----ProTyrLeuIleArgGlyTyrPro--- 1674
Db 3942 ---TATCCGTCGCTCCTCGTGAAGCTCGGCGGTACTCCGTCGCGTGGCGCGCAGC 3886
QY 1675 AspThrAlaAlaLeuGluAsnArgGlnThrIleAsnAspTyrIleThrSerGlnGln 1694
Db 3885 AACATCTCCGCGTCCGCGCGCGAGTACTCGGGGAGCACTCCGCGCTCCTCGAGGGCC 3826
QY 1695 MetHisAsnThrAlaThr-----AlaMetAlaGlnArgAla 1707
Db 3825 TCCACAGACCTCCAGCAGCAGCGCTGTGGGGTCCATGCGCGTGGCGCTCGCGGCG 3766
QY 1708 AspMetLeuArgGlyLeuSerProArgGlySerLeuAlaLeuAsnTyrAlaAlaGly 1727
Db 3765 GATATC-----CCGAAGACTCC-----GCGTCGAACCTCGCGCGC--- 3730
QY 1728 ProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProThr 1747
Db 3729 -----GAGTGCAGGAACCCCGCTCCCGCACATAGGTCTTGGCGGCGTC 3685
QY 1748 ProGly-----ThrProAlaThrAlaMetAspArgLeuAla-----Tyr 1760
Db 3684 CCGGCTCGGGTCTGACACCGCTCGAGTCCAGCGCGGTCTGGTGGGAAGGACGAG 3625
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSerProGly 1780
Db 3624 ATGGCGTCCGCGCGGACTCCAGCAGCTCCAGAGAGCGCGCGAGTGCAGCGCGCGC 3565
QY 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerGluArgGluArgAspArg 1800
Db 3564 GGGAGCGCGCAGCGGACA---CCACAGATGCGGATCGGTCTCGTCCCGGCTGCCG 3508
QY 1801 AsArgGluArgAspArgGluArgGluArgGluArgGluArgGluArgGluArgGlu 1820
Db 3507 CCGCGGCGTCCGACACCTTACCGCGCGCAGAGAGCACTCCGGCGCCACCGGCAAC 3448
QY 1821 ValGluHisAlaPro-----IleTrpArgProGlyThrGluGlnSerSerGly 1836
Db 3447 GAGTCTTCGCGCGCGCGGTCTCCACGATGTCTCGCGCAGCGCAGCGCGGTGGC 3388
QY 1837 SerSerGlySerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHis 1856

Db 3387 AGTCTCTAGACCAAGCGTGGCGGCGAGGGGAGGCCCGAGGCGCCCTTACGACGGTGCAC 3328
QY 1857 -AlaHisGlnHisSerProIleSerPro-----ArgThrG1 1868
Db 3327 AGTCTACCGCGGTACCG---AGTCCATCCCTGGGAGGTGAAGTTCGGTTCGAGCGCG 3271
QY 1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleLeu 1888
Db 3270 ACCTCCCGGATCGT--- 3252
QY 1888 xAlaValGluProSerLysProThrValLeuArgSer-----ThrSe 1902
Db 3251 -----GGCCCCAGACGCGCGGTCTCTCCGACACAGCGGAGCGATCTCGGG 3199
QY 1902 rThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuG1 1922
Db 3198 GACATCGCTCCGTCGACCGGAGT-----CCGCTCGTCCGTCCTCGTGG 3148
QY 1922 yGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAl 1942
Db 3147 GACGTCC-----CCTCGTGTCTC 3130
QY 1942 aProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAla 1962
Db 3129 TCCGCTCCGGGAGAGCGTGTCCCAAG-----CATCGTGGCGGTGACA 3082
QY 1962 sProProAlaArg-----SerGlyLeu----- 1969
Db 3081 CCGCGCCCGCGTCCGCGGTGACCGCGCGCTCAGCGCGGCGACCATGGTACGCGT 3022
QY 1969 ----- 1969
Db 3021 TGAAGGGTGTGGGAGTTCGATCGCGCGCGGTCTCCCGTGTGGAGTGGCGTG 2962
QY 1970 -GluProAlaSerSerProSerLysGlySerGluProArgProGluProValProValSe 1989
Db 2961 GCGTCCAGCGCGCGTCCCGGTGAACCGTCCGCGCGCGCTCGAACACCTCCCGC 2902
QY 1989 rGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPr 2009
Db 2901 TCCGACGCGCC---GCCGATCGCGGACCA-----CCACCATCATCACCGCC 2851
QY 2009 o-----AspProProAlaProProAlaSerAlaSerAsp----- 2020
Db 2850 CCAGGCACTCACCGCATCCCGTCAGCACCCCATCGGAGCCACCTCCACCAACGTC 2791
QY 2021 -ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluArgSe 2040
Db 2790 CCGACACCGACGACGACCCACCCCA-----TCGGGAAGCGACCGCTCCCGCGC 2732
QY 2040 r-----LeuGlyTyrHisGlySerTyrSerProGluGlyValGluProValSerPr 2058
Db 2731 ATGACGACCCCAATACCGGATCCCGCACTCCCG-----GAACCCACCAACCC 2681
QY 2058 oValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAs 2078
Db 2680 ACCGACACACCGGACACCCACCCACCGCGCG-----CACCGGACCACTCCA 2628
QY 2078 pLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysGlyG1 2098
Db 2627 ACACCTCCAC-----AACCCCGGAACTCCCCCA 2598
QY 2098 yGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerPr 2118
Db 2597 ACACCGCTCCATCAACACCGAATGAACCCATCGACACATCAACCG-----CC 2547
QY 2118 oLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLeuAlaGlnHi 2138
Db 2546 GGCACCTCCACCCCATCCCATCAGGAGGACCAACCGACTC-----CAGCA 2499
QY 2138 sIleSerGluValIleThrGlnAspTyrThrArgHisHisPro-----GlnGlnLeuSe 2156
Db 2498 CACCCAC-----ATCACCGCAACACCAACCGCGCGCCCATTCACCGCGCAACCG 2445

| | | | | | | | |
|----|------|--|------|----|------|---|------|
| Qy | 922 | SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg | 941 | Qy | 1226 | gGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIle | 1246 |
| Db | 6165 | CGGCGTCCCGGTGAACACCGTCGCGAGCGCGCTCGAAACACCTCCCGCTCCGACCG | 6106 | Db | 5176 | CCCCATACCAACCACTGGCACCTGACCCGTG-----AACACACCCCAACC----- | 5128 |
| Qy | 942 | ProSerLeuLeu-----ThrProThrGlyAspProArgAlaAsnAlaSerProGln | 958 | Qy | 1246 | eileGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGle | 1266 |
| Db | 6105 | CCCGCGCATCGCGCGACCAACCAACCATCATCACCAGCCGCCCGGAGGCACTCACC | 6047 | Db | 5127 | -----CCACCCGCGACAGCCCGCGACCGACACCCCGACACCTCCCGGCGAGC | 5075 |
| Qy | 959 | LysPro-----LeuAspLeuLysGlnLeuLysGlnArgAlaAla | 972 | Qy | 1266 | yHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerVa | 1286 |
| Db | 6046 | CATCCCCGTGACACCCATCGCGAGCCACCTCCACCAAGTCCCGACACCCAGACCG | 5987 | Db | 5074 | AAAC----- | 5071 |
| Qy | 973 | IleProIleGlnValThrLysValHisGlnProProArgGluAspAla-----Ala | 990 | Qy | 1286 | lThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAl | 1306 |
| Db | 5986 | CACCACCCCAACCCATCCCGGAACGACCG-CTCCCGCATGACGACCAACCAATACC | 5928 | Db | 5070 | -----CCACCAAGACGACGAGCTCCCGCCCGCCACACCGGACCGCACCGC | 5021 |
| Qy | 991 | ProThrLysProAlaProProAlaPro-----ProProPro | 1002 | Qy | 1306 | aProLysArgThrTyrAspMetGluGlyArgValGlyArgAlaIleSerSerAlase | 1326 |
| Db | 5927 | CCGATCCCCCACTCCCGGAACCCACCCACCCCGGACACACCCCGGAACTCCGCCA | 5868 | Db | 5020 | A-----CGGTGCTCGAAAGAGTCCGCGTCGAAGC | 4991 |
| Qy | 1003 | GlnAsnLeuGlnProGlu-----SerAspAlaProGlnProGlnProGlySerProArg | 1020 | Qy | 1326 | rileGluGlyLeuMetGlyArgAlaIleProPro----- | 1337 |
| Db | 5867 | CACCGCGCGCACCGGACCGAACTCCACGACTCCACACCCCGGAACTCCGCCA--- | 5811 | Db | 4990 | CAGCGAATAAGCCACATCCCGGATCCGCACCCCGCTCCACTACTACCCCAACCG | 4931 |
| Qy | 1021 | GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln | 1040 | Qy | 1338 | -----GluArgHisSerProHisLeuLysGluGlnHisIleArgGlySe | 1354 |
| Db | 5810 | -----ACACCGGCTCCATCAACACCGGAATGAACCCCATCGC-----ACACAT | 5769 | Db | 4930 | CACCGCTCGCGCGCAACGCCCTCCGACCGCGGACAAACACACGCGGCGGACGAGC | 4871 |
| Qy | 1041 | LysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProValProProArg | 1060 | Qy | 1354 | rileThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGle | 1374 |
| Db | 5768 | CCAACCGCGGACCTCCACCCATCCCGCATCCCGGAGGACCAACCGACTCCA----- | 5715 | Db | 4870 | AGCGCGTCCGGTACCCTCCAGGACGCGGTC-----GATGTCTCGACCCACT | 4812 |
| Qy | 1061 | GluValIleLysAlaSerProHisAlaPro---AspProSerAlaPheSerTyrAlaPro | 1079 | Qy | 1374 | uAlaLysLeuLeu-----LysArgGluGle | 1382 |
| Db | 5714 | -----GCACACCCACATCACCCGACAAACACCAACCGAGCGCGCCATCCACCG | 5667 | Db | 4811 | CCGCGCTCTCGATCATCAGATCGCGTCTGCCACTGATCCCGAACGACGACACAG | 4752 |
| Qy | 1080 | Pro-Gly---HisProLeuProLeuGlyLeu-----HisAspThrAlaAr | 1093 | Qy | 1382 | yThrPro-----ProProProProSerArg-----As | 1392 |
| Db | 5666 | CGCAACCGACACCCACTCCCGCAACCCCTCAACACCCCGGACCAACCGGCGCCATCC | 5607 | Db | 4751 | CCGCGCGCGCGCGCTCCGACCGCGGCGCACCGCGCGCTCTGCCAGCGGCGGA | 4692 |
| Qy | 1093 | gProValLeuProArgProThrIleSerAsnPro----- | 1105 | Qy | 1392 | pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGle | 1412 |
| Db | 5606 | CCCCCACCAGCACATCCCCCACCACCGCGCAACCCATCAACCCCGCGGCGCA | 5547 | Db | 4691 | CGTGCCCGACGACAGTCAGTCGCTGGCGCGTCCGCGTGAAGTGTCCGCGCA | 4632 |
| Qy | 1106 | -----ProProLeuLysSerAlaLysHisProSerValLeuGluArgGlnIleGle | 1123 | Qy | 1412 | uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGle | 1432 |
| Db | 5546 | CCACCAACCGGACCGCATCACCAACGACACACCCCGCGCATACACGACGACG---- | 5492 | Db | 4631 | CGGTGCGCTGCTCCAGGGCAAG-----CAGCA | 4605 |
| Qy | 1123 | yalalleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAl | 1143 | Qy | 1432 | uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh | 1452 |
| Db | 5491 | -----CACCTCCCGCCACCG-AATGACCCCAACACC | 5464 | Db | 4604 | TCTTGACCAACAC-----CGCGCGCGCGC-----GCGCGCTG-- | 4571 |
| Qy | 1143 | aProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaPr | 1163 | Qy | 1452 | xGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAs | 1472 |
| Db | 5463 | ACCGACACCTCCA-----CACCCGAGCTTCAACCGCGGCAACAAACCGCCACC | 5416 | Db | 4570 | -----GGTGTGACCGACGTTGGACTTCAGCGACCCGA | 4539 |
| Qy | 1163 | oPheSerGlyValLysGlnLeuSerProArgGlyGlnAlaGlyPro----- | 1180 | Qy | 1472 | pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa | 1492 |
| Db | 5415 | TCCACGCAACACGCGAGGTGAGAACTCCGTCGCGACCCCAAC-ACCCACCCACAGA | 5357 | Db | 4538 | GCCACAAGGGGTC-----GCCACCGCGCTC----- | 4514 |
| Qy | 1181 | -----ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh | 1198 | Qy | 1492 | lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgPr | 1512 |
| Db | 5356 | ACCGNACCAAGAACCGACCCAGACCGCGGACCGCATCAGCCCGGCGGACCGCG | 5297 | Db | 4513 | -----ACGCGCGTACGTCGCGACGCGCTCGCTCGATCGGATCGCCCGCGCTCC | 4458 |
| Qy | 1198 | raLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPr | 1218 | Qy | 1512 | oGlyThrAla----- | 1516 |
| Db | 5296 | GTCCAGCTCGACGTCCCGAATCATCATCCCGCAACGACGACCATCCACCTCCCGCAC | 5237 | Db | 4457 | CCGTACCGTGGCGCTCCACCGCGTCCACGTCCCGCGGCGAGCGCGCTCGGCCAGGG | 4398 |
| Qy | 1218 | oSerAspSer-----AlaIleThrTyrAr | 1226 | Qy | 1516 | rSerSerGlyGlySerIleAlaArgGlyAla----- | 1527 |
| Db | 5236 | CATCGAACACCTCATCCAGCACCTCCGCAACACCCCAACCCCGCGCATACACCCAG | 5177 | Db | 4397 | CAGCTCGGATGACCCCTCTCGCGGAGCGGTTGGAGCGGTCAGCGCGTGTGTGGCAC | 4338 |
| Qy | | | | Qy | 1527 | oValIleVal---ProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHis | 1546 |

FT /transl except= (pos:350..352, aa:Met)
 FT /note= "ORF1 encodes protein shown in AAW23716"
 FT CDS 14046..20036
 FT /*tag= b
 FT /note= "ORF2 encodes protein shown in AAW23717"
 FT CDS 20110..31284
 FT /*tag= c
 FT /note= "ORF3 encodes protein shown in AAW23718"
 FT CDS 31329..36071
 FT /*tag= d
 FT /note= "ORF4 encodes protein shown in AAW23719"
 FT CDS 36155..41830
 FT /*tag= e
 FT /note= "ORF5 encodes protein shown in AAW23720"
 FT
 XX EP791656-A2.
 PN
 XX
 XX 27-AUG-1997.
 PD
 XX 19-FEB-1997; 97EP-00301066.
 XX
 XX 22-FEB-1996; 96US-0012050P.
 XX
 XX (ELIL) LILLY & CO ELI.
 XX
 XX Burgett SG, Kuhstoss SA, Rao RN, Richardson MA, Rosteck PR;
 PI
 XX WPI; 1997-418047/39.
 XX
 DR P-PSDB; AAW23716, AAW23717, AAW23718, AAW23719, AAW23720.
 DR
 XX
 XX DNA encoding Streptomyces ambofaciens platenolide synthase domain - for
 PT production of spiramycin-related polyketide antibiotics.
 PT
 XX
 XX Claim 9; Page 8-33; 81pp; English.
 PS
 XX
 CC This sequence represents the platenolide synthase gene cluster of the
 CC invention. This sequence is referred to as the smg gene, and was
 CC isolated from Streptomyces ambofaciens. This sequence encodes the multi-
 CC functional proteins which direct the synthesis of the polyketide
 CC platenolide, platenolide. The DNA can be used to produce compounds
 CC antibiotic spiramycin. The DNA can be used to produce compounds
 CC exhibiting antibiotic activity based on the platenolide structure,
 CC including specifically the macrolide antibiotic spiramycin and spiramycin
 CC analogues and derivatives. Modifications of the platenolide synthase DNA
 CC sequence can be made so as to change the number and type of carboxylic
 CC acids incorporated into the growing polyketide chain and to change the
 CC kind of post-condensation processing that is conducted
 CC
 XX
 SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.39e-11 Length: 44377
 Score: 670.00 Matches: 665
 Percent Similarity: 31.22% Conservative: 282
 Best Local Similarity: 21.93% Mismatches: 1121
 Query Match: 5.07% Indels: 980
 DB: 2 Gaps: 149
 US-09-522-753-5 (1-2517) x AAT78508 (1-44377)
 QY 9 AlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProVal 28
 Db 9254 GCCCGTACGTGGCCAGCAGCGCCCGCTCGATCGATCACCAGCGCGTCCCGGTAC 9195
 QY 29 GlnIleAlaAlaThrHisThrAspValGlyLeuLeuGluTyrGlnHisSerArgAsp 48
 Db 9194 CGTGGCGCTCCACCACAT----- 9177
 QY 49 TyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgProSerLeu 68
 Db 9176 -----CCAGTGGCGGGGACCAGCCCGG 9153

QY 69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSer 88
 Db 9152 CATCAGCCAGCGCTCAGGATCACCCTGCTGCGCGGACCATCGCGCGCTCAGCC 9093
 QY 89 HisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgPro 108
 Db 9092 CATTGC-----TCGACCGCTCT 9075
 QY 109 ArgLeuGluLeuLeuProAsp-----ProLeuLeuArgProSer 121
 Db 9074 GA-TTCACCGCACTGCCCGCAGCACCAGCCAGCAGCCGATGACCTGCGGCGCGCTCG 9016
 QY 122 ProLeuLeuAlaThrGlyGln---ProAlaGlySerGluAspLeuThrLysAspArgSer 140
 Db 9015 GAAAGCGCTCCACCAGCAGCACACACACCTCCGCCACCGCTCCGCTC-CGCACC 8957
 QY 141 LeuThrGlyLysLeuGluProValSerProProSerProPro----- 154
 Db 8956 CTCGCGAAGCGCTTGCACCGCTCCGCCCGCCAGCCCGCTCGCGGAGAACTCCAC 8897
 QY 155 -----HisThrAspPro-----GluLeuGluLeuVal 163
 Db 8896 GAACACCGTCGGCGTCCCATCACCCTACCCAGCCGCAACGCGCAGATCAGCTCACC 8837
 QY 164 ProProArgLeuSerLysGluLeuIleGlnAsnMetAspArgValAspArgGluIle 183
 Db 8836 CCGCGC-----CAACGACTGCACCGCCAGATGCAACGCGCAGCAGCAGCAG 8789
 QY 184 ThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGlu---Glu 202
 Db 8788 ACA-----CGCCGTGTCCACCGTCCAGCGCGCGCTCCCAACCCCA 8747
 QY 203 GluAlaAlaLysPro-ProGlu-----ProGluLysProValSer-----ProPr 217
 Db 8746 CGCATAGGCAACCGACCGAGATCACCTGCGAGGCTCCCGTCCAGCAGATGCGCGCC 8687
 QY 217 oProlleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 237
 Db 8686 GTAGGTGTCGGCGCGTCCCGCAGCTGGGTGCG-GTAGTCTCTCTGGAGATCCCGCTGA 8628
 QY 237 s-----AlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLe 254
 Db 8627 AGACCGGTCCGGTCCCGCGCAGCGAGTCCGGAACGATGCGCGCGCTCCAGCGCT 8568
 QY 254 uPro-----LeuTyrAsnGlnPro-----SerAspThrArgG 265
 Db 8567 CCCACGACGTCTCCAGCAGCAACCGTGTCTCGCGGTCCATGCGCGCTCGCGCGCG 8508
 QY 265 nTyrHisGluAsnIleLysIle-----AsnGlnAlaMetArgLysLysLeuIleLeuTy 283
 Db 8507 ATATCCCGAAGAACTCCCGCTCGNACTCGCGCGGGAACGCGAGAACCCCGCTCCCGTA 8448
 QY 283 rPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAs 303
 Db 8447 CATAGCTTTGCCCGCAGCACCCTCG-----CTCCGGTCTGTACAGCCCTCGA 8400
 QY 303 pGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsnProArgArgAr 323
 Db 8399 CGTCCACGCCACGTTCGGCGCGGAACGGGAGAGCGGTGTGCCCTCACCAGCCAC----- 8345
 QY 323 gAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysG 343
 Db 8344 -----CAGGTCCACAGGTGCTCCGCGC 8322
 QY 343 nArgLeuGlnGluArgMetGlnSerArg----- 353
 Db 8321 ACGCCAC---ACCGCGCGGTAAACGCGCAGCCATGCCCATCGGATCGGCTACCGCT 8265
 QY 354 valGly-----GlnArgGlySerGly-----LeuSerMetSerAlaAl 366
 Db 8264 CGTCGGCATACCGCGGTGGGGCGCGCTGCGGCTCCGCTCGCTGCTGGGACCGA 8205
 QY 366 aArg-----SerGluHisGluValSerGluIleIleAspGlyLeuSe 380

1969 QY ----- 1969
3021 TGAAGGGTAGTGGAGGTGATGTCGCGCGCGCGTGTCTCCGGTGTGGAGTGCCTG 2962
1970 -GluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSe 1989
2961 GCCTCAGGCGCGGTCTCCGGGTGAACACCGTGCAGCGCCGCTCGAACACCTCCCGC 2902
1989 rGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPr 2009
2901 TCCGACGGCCCGC---GCCCATCGCGGCACCA-----CCACCACATCATCACCGCC 2851
2009 o-----AspProProAlaProProAlaSerAlaSerAsp----- 2020
2850 CCGAGCCTCAGCCGCGCATCTCCCGTCAGCACCCCATCGGAGCCACCTCCACCAACGTC 2791
2021 -ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuLysArgSe 2040
2790 CCGACACCCAGACCGCACACCCCGCCCA-TCCGCGAAAGCGACCGCCTCCCGCGC 2732
2040 r-----LeuGlyTyHisGlySerSerTySerProGluGlyValGluProValSerPr 2058
2731 ATGACGACCCATATCCCGGATCCCCCACTCCCG-----GAACCCACACCCCG 2681
2058 oValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAs 2078
2680 ACCGACACACCGGACACACACACCGCGCG-----CACCGACCGAACTCCA 2628
2078 pLysSerHisLeuGluGlyLeuLysArgProLysGlnProGlyProValLysLeuGly 2098
2627 ACAGCTCCAC-----AAGCCCGCGAACTCCCGCA 2598
2098 yGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerPr 2118
2597 ACACCGCTCCATCAACCGAATGAACCATCGGACACATCCACCG-----CC 2547
2118 oLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLeuAlaGlnHi 2138
2546 GGCATCCACCCATCCCGGATCCAGCGAGGCAACCCGACTC-----CAGCA 2499
2138 sIleSerGluValIleThrGlnAspTyThrArgHisHisPro-----GlnGlnLeuSe 2156
2498 CACCCAC-----ATCACCGCAACACACCGCGCGCGCCATTCACCGCGCGCAACCG 2445
2156 rAlaProLeuProAlaProLeuTySerPheProGlyAlaSerCysProValLeuAspLe 2176
2444 ACACCCACTCCCGCAACCCCTCAACACCCCGCGCACCGCA-----CT 2400
2176 uArgArgProProSerAspLeuTyLeuPro-----ProProAspHis----- 2190
2399 CCGACGCGCC-----CACCGACACATCTCCCGCCACCGCACCGCAACCCACCCCA 2352
2191 ----GlyAlaProAlaArgGlySerProHisHisSerGluGlyLysArgSerProGluPr 2209
2351 TCACCCCGCGCGCGCACCAACCGCGCGCGCATCACCAACGACCAACCCCGCGCA 2292
2209 oAsnLysThrSerValLeuGlyGlyGlyAlaAspGlyIleGluProValSerProGl 2229
2291 CATACGAGCGACCATC----- 2268
2229 uGlyMetThrGluProGlyHisArgSerAlaValTyProLeuLeu----- 2245
2267 CCGAATGACCGAA-----CACACCGACACCTCCACACCCCGCGCTCCCAACGCGCGGA 2214
2245 ----- 2245
2213 ACAACGCCACTCAACGCAAAACACGACGGTGAGCAAACTCGTCCGACCCCAACCAACC 2154
2246 ----TyArgAspGlyGluGlnThrGluProSer---ArgMetGlySerLys-Ser---- 2261
2153 CACCCACAGAACCGAACCGAACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2094
2262 -----ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnS 2280

2093 CGGCACCCCGCTCCACGTCGACGTCGCG-----AACATCATCCTCCCGCAACG 2046
2280 erAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAnThrHisAsnArgA 2300
2045 ACGGACCATCCACCTCCCGCCACCATCGACACACCTCATCCAGCAGCTCCGCAACACCC 1986
2300 snGluProGluTyrAsnIleSerGlnProGlyThrGlu-----IlePheA 2315
1985 CACCCCGCATACAAACCCACGCGCCCATACCAACCCACTCGGCACCTGACCGGTGAACA 1926
2315 snMetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluH 2335
1925 ACACCCCGCCACCCACCGCGCACAGCC----- 1900
2335 isAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyra 2355
1900 ----- 1900
2355 spGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerA 2375
1899 --CGCCGCGAGCGCACACCCCGCACACCTCCCGCGGATCCCGCGCCACCTCCAGTACTC 1842
2375 laSerLeuPro----- 2378
1841 CGAGCTCCCGCCCGCCCGCCACCGCACCGCACCGCGTCTCGAAAGAGTCCGCGTCG 1782
2379 --AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProG 2398
1781 AGCCAGCGAATAGCCACATCCCGCGGATCCCGCGCCACCTCCAGTACTCAGCCCA 1722
2398 lyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA 2418
1721 ACCGACCGCTCGCGCGCG-----AACGCGCCCTCCGACCGC---GCGGACAAACA 1671
2418 laProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyA 2438
1670 GCGCGCGGACGTCGGAAGCGAGCGCGCGACCGCGCGCGCGCGCGCGCGCGCGGCA 1611
2438 spCysAsnArgArgThr-----ProLeuThrAsnArgValTrpGluAspA 2453
1610 GGTGTTCAGGACCATCGCATGTCGTCGACCGATGCGGAC-----GAACTGA 1560
2453 rgProSerSerAlaGlySerThrProPheProTyArgProLeuIleMetArgLeuGlnA 2473
1559 CACCGCGGACGAGGGGTGCGCGCC----- 1534
2473 laGlyValMetAlaSerProProProProProProProProProProProProProPro 2491
1533 --GGGCGGAGCGGACGCTCGCGCGGACGGTGTGCACCTTCAGCGGAGCTGTTCAGCG 1476
2492 AlaGlyProHisAlaTrp 2498
1475 GGATGCGCGGAGAGGGGTGG 1455
RESULT 33
AAT78508/c
ID AAT78508 standard; DNA; 44377 BP.
XX
AC AAT78508;
XX
XX 26-FEB-1998 (first entry)
DT
XX
DE Platenolide synthase gene cluster.
XX
KW Platenolide synthase gene cluster; platenolide production; srmG gene;
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS Streptomyces ambofaciens.
XX
PH Location/Qualifiers
FT Key 350..14002
FT CDS /*tag= a

Db 4870 AGCGCGTCCGCGTACCGCGGTACGAGCAGCGCGCTC-GATGTCTCGACCCACT 4812
 Qy 1374 uAlaLysLeuLeu-----LysArgGluG1 1382
 Db 4811 CCGCGCGCTCTCGATGATCAGATCGCGGTTCGTCCACATGATCCCGAACGACGACAG 4752
 Qy 1382 yThrPro-----ProProProProProSerArg-----As 1392
 Db 4751 CCGCGCGCGCGCGGTCCGACCGCGCGGCGCCACCGCGCGCTCTGCCAGCAGGCGGA 4692
 Qy 1392 pLeuThrGluAlaTyrlYsThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisG1 1412
 Db 4691 CGGTGCCCGACGACGACGTGCGTGTGGCGCGTCCGCGGTGAGTGTCCGCGGA 4632
 Qy 1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluLeuProArgGluG1 1432
 Db 4631 GCGTGGCGTGTCCAGGCGAAG-----CAGCA 4605
 Qy 1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh 1452
 Db 4604 TCTTACACACACC-----CGCGCGCGCGC-----GGCGCGCTG-- 4571
 Qy 1452 rGlnGlyThrProLeuLysTyAspThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
 Db 4570 -----GGTGTACCGACGTGTGACTTCAGCGACCCGA 4539
 Qy 1472 pValArgSerLeuIleGlySerProGlyArgThrPheProValHisProLeuAspVa 1492
 Db 4538 GCCACAGGGGTC-----GCCACCGCGTC----- 4514
 Qy 1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyGluGluSerLeuLysSerArgPr 1512
 Db 4513 -----ACGCGCGTACGTGCGCAGCGCGCTCGCTCGATCGATCGCCAGCGCGCTCC 4458
 Qy 1512 oGlyThrAla-----Se 1516
 Db 4457 CCGTACCGTGGCGCTCCACCGGTCCAGCTCAGCCGCGGAGGCGCGGTCCGCGCAGGG 4398
 Qy 1516 rSerSerGlyGlySerIleAlaArgGlyAla-----Pr 1527
 Db 4397 CAGCTCGATGACGCGCTCTCGCGGAGCGGTGGGAGCGGTGAGCGGTGTCGGCAC 4338
 Qy 1527 oValIleVal--ProGluLeuGlyLysProArgGlnSerProLeuThrTyGluAspHis 1546
 Db 4337 CGTCTGTGTGACCGCATCGCCACCGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 4296
 Qy 1547 GlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArg----- 1563
 Db 4295 GGTGTCCCGCACGCGCGGT---CGGAGAGCGGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 4239
 Qy 1564 -----GluProThrProArgLeuGlnGlySerLeuSerSerLysAlaSerGln 1581
 Db 4238 CGGACCGAGCCGTACCGTCCGATCGCGGAGACGGATCGAGCGCGGTGCGGGGCGA 4179
 Qy 1582 AspArgLysLeuThrSerThrProArg-----GluIleAlaLysSerProHisSerThr 1599
 Db 4178 GCCCGCGTCCGCGGAGACTCCAGAACATCGCGCGCGCGCATCATCGTGGCGCGC 4119
 Qy 1600 ValProGluHisHisProHisProIleSerPro----- 1610
 Db 4118 CGGCCAGAGCC-AGCCCGCAC-----TCGCGCGCGCGCGCGCTGACCGCGCGGTGC 4066
 Qy 1611 -----TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyArg 1625
 Db 4065 AGCGCCACCGAGGACGAGGAGACCGCGGTGTGACGCGTACGCGTACG----- 4024
 Qy 1626 SerHisIleProLeuAlaPheAspProThrSerIleProArg-----GlyIleProLeu 1643
 Db 4023 -----GCCGTGCGCG-----GTGCCAGGGGTGTAGCGATTCGCGCG 3985
 Qy 1644 AspAlaAlaAlaAlaTyrlYrLeuProArgHisLeuAlaProAsnProThrTyrlYrHis 1663
 Db 3984 GAGGCCACGCTCGCG-----GTGTGCGGTGAGCAGGTACCCCTCG 3943

Qy 1664 LeuTyrlPro-----ProTyrlLeuIleArgGlyTyrlPro--- 1674
 Db 3942 ---TATCTCGTCCGTCCCTCGTGAAGCTCGCGCGTACTCCGGTGGTGGCGCGACG 3886
 Qy 1675 AspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrlIleThrSerGlnGln 1694
 Db 3885 AACACTCCGGCGTCCGCCCGCGCAGTACTCGGGAGCAGCTCCGGCGTCTCTCGAGGGCC 3826
 Qy 1695 MetHisHisAsnThrAlaThr-----AlaMetAlaGlnAla 1707
 Db 3825 TCCACAGCGCTCCAGCAGCAAGCGCTGCTCGGCTCCATGGCGTGGCTCGCGCGC 3766
 Qy 1708 AspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrlAlaAlaGly 1727
 Db 3765 GATATC-----CCGAAGAACTCC-----GCGTCAACTCGCGCGC--- 3730
 Qy 1728 ProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThr 1747
 Db 3729 -----GAGTGCAGGAACCGCGCTCCCGCACATAGGTCTTTCGCGCGCTC 3685
 Qy 1748 ProGly-----ThrProAlaThrAlaMetAspArgLeuAla-----Tyr 1760
 Db 3684 CCGGCTCGGGTGTGTACAGCCCGTCCGAGGTCCAGCGCGCTCGGTGGGGAAGACGAG 3625
 Qy 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
 Db 3624 ATGGCGTCCGCGCGCGACTCCAGCAGTCCAGAGAGCGCGCGAGTCCAGCGCGCGC 3565
 Qy 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg 1800
 Db 3564 GGGAGCGCGACGCCGACA---CCACAGATGGCGATCGGTGCTGCTGCTCCGCGTCCG 3508
 Qy 1801 AspArgGluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSerThrThr 1820
 Db 3507 CCGCGGCTCCGACCCCTACCGCGCGCGAGAGCAGTCCCGCGCCACCGCGCAAC 3448
 Qy 1821 ValGluHisAlaPro-----IleTyrArgProGlyThrGluGlnSerSerGly 1836
 Db 3447 GAGTCTTCGCGCGCGCGCTCCACGATGTGTCCGCGCAGCGCACGCGGGTGGGC 3388
 Qy 1837 SerSerClySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHis 1856
 Db 3387 AGTCTGAGACCGAGC 3328
 Qy 1857 -AlaHisGlnHisSerProIleSerPro-----ArgThrG1 1868
 Db 3327 AGTCTGACCGCGGTCAACG---AGTCCATGCTCCGCGAGGTGAAGGTGCGGTGCGCGC 3271
 Qy 1868 nAspAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
 Db 3270 ACCTCGCGCGGATCGTGT----- 3252
 Qy 1888 rAlaValGluProSerLysProThrValLeuArgSer-----ThrSe 1902
 Db 3251 -----GGCCACGAGCGCGCGTCTCCGCGACCGAGCGCGCATCTCGGGG 3199
 Qy 1902 rThrSerSerProValArgProAlaThrPheProAlaThrHisCysProLeuG1 1922
 Db 3198 GACATCGGTCTCCGTGACCGCGAGT-----CCGCTCGTCCGCGTCCCTGTGG 3148
 Qy 1922 yGlyThrLeuAspGlyValTyrlProThrLeuMetGluProValLeuLeuProLysGluAl 1942
 Db 3147 GACGTCC-----CCTCGGTGCTC 3130
 Qy 1942 apoArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLy 1962
 Db 3129 TCCGGTCCGGGAGACGCTGTGCCCCACCG-----CATCGGTGGCGGTGACA 3082
 Qy 1962 sProProAlaArg-----SerGlyLeu----- 1969
 Db 3081 CCGCGCGCGGTGCGCGGTGACCGCGCGGTGACCGCGCGGTGAGTGGGTACGCGGT 3022

Db 16188 CGCCACCTA-----CGG 16199
Qy 1381 GluGlyThrProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db 16200 ACAGGACCGCCGGCAACAAACCGCTGGTGGCTCATGATCGACATCGGCCA 16259
Qy 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGlu----- 1412
Db 16260 CGCGCAGGCTCGCGCAGGTGTGGCGGGGTCAACAAGATGGTGTGATGGCGGTGCGGAATGG 16319
Qy 1413 -----GlyLeuVal----- 1415
Db 16320 TCTGCTCGCGGACGTTGCATGTGATGATGACCGCTCGCCGATGTGTGATGTGCGCGGG 16379
Qy 1416 -----AlaThrValLysGluAla 1421
Db 16380 GCGGTGTCAGCTGTCAGCGAGACGGTGCCTCGCCCGGGGGGGGGGGTGGCGG 16439
Qy 1422 GlyArgSerIle-----HisGluIleProArgGluGluLeuArgHis 1435
Db 16440 GGCAGGAGTGTCTCGTTCGGCGTCAGCGGCACCAACCGCTCATCGAAG-AAGCAC 16498
Qy 1436 ThrProGluLeu-----ProLeuAlaProArgProLeuLysGlu 1448
Db 16499 CGCCCCCAACATCATCCCTCAGACACACCCCGCAGACGCGCCCGGAGAGACGCG 16558
Qy 1449 GlySerIleThrGlnGly-----ThrProLeuLysTyrAspThrGlyVala 1463
Db 16559 CGACGAGTGTGGGGGAGAGCGCGCGCGCAGCAGCGCGT-----ACGCGGGG 16608
Qy 1464 SerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGly----- 1481
Db 16609 GAAGCGACTGCT-----CCTGCTCGCGGAGTCCAGCGGTGTGG 16647
Qy 1482 -----ArgThrPheProProVal----- 1487
Db 16648 CCCTGGCTGTGTGCGCAAGTGTGACGCGCGCTGCGCGCCCGAGCGCCCTGTGCAC 16707
Qy 1488 -----HisPro-----LeuAspVal-----MetAlaAsp 1495
Db 16708 GCCCACTCACGACACACCGCGCTGCTGCTGCGCGACGTGCGGTACACCTCGCCAC 16767
Qy 1496 AlaArgAlaLeu-----GluArgAlaCysTyrGlu 1505
Db 16768 GCGCGCGCTGTGTGACACCGCGCCACCTCATCGCGCGCGACGCGACACTTCTCTG 16827
Qy 1506 GluSerLeuLysSer-ArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgG1 1525
Db 16828 CAAGCACTCCAGGACACTCGCGCGAGGCGAACCACCCCGCGCTCATCCACAGCA----- 16882
Qy 1525 YAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAs 1545
Db 16883 -GGCCCCAGCGGGGACCGGGACCGGGGAGCGCGCAGGAAGACCGCATTCATCTGTCTCC 16941
Qy 1545 pHisGlyAlaProPheAlaGlyHisLeuProArg----- 1556
Db 16942 GGACAGGGGACCCCAACCGCGCGCATGGCGCACCGGCTTACACACACCCCGCTTTC 17001
Qy 1557 -----GlySe 1558
Db 17002 GCGCGCGCATCTAACGACATCTGACCCCATCTGACCCCGCCACCTCGACACCCCGCTCTC 17061
Qy 1558 rProValThrMetArgGluProThrProArgLeu----- 1569
Db 17062 CCCCTCTCACCGAGGCCCAACACCCAGACACACACACCTCTCGAAGAGCGGCCGCA 17121
Qy 1570 -----GlnGluGlySerLeuSerSerLys 1578
Db 17122 CTGCTCCAGCAGACCCCGTAGCCAGCCCGCGCTTCTGCGCTTCCAGGTGCGCCTCCAC 17181
Qy 1578 sAlaSerGlnAspArgLysLeuThrSerThrProArgGluIle-----Al 1593
Db 17182 CGCCTCTCACCGGCGGTACACATACACCCCGCCACTACTACGCGCGGACACTCCCTCGGC 17241

Qy 1593 alysserPro-----HisSerThrValPro-GluHisHisProHisP 1607
Db 17242 GAAATCACCGCGCCACCTCGCGCGCATCTCCTCACCTCACGACGCCACCTCATC 17301
Qy 1607 rolleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerH 1627
Db 17302 ACCAACCGCGCCACCTCATGCAACCATGCCCCCGCGCACCATACCCCTCCACACC 17361
Qy 1627 istleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAla 1647
Db 17362 ACCCCCAACCATCATCACCCACCTCATCGGCCCAAGAACCTCTGCGCATCGCGCC 17421
Qy 1647 laAlaTyr-----TyrLeuProArgHisLeuAla----- 1656
Db 17422 ATCAACACCCCGACCTCTCTGTCATCAGCGGCACCCCGCACCTCGTCAACACATCAC 17481
Qy 1656 ----- 1656
Db 17482 ACCCTCTGCCAACAAACAGGCATCAAAACCAACCTCCCCACCAAAAGCGCTTCCAC 17541
Qy 1657 -----ProAsnProThr-----TyrProHisLeu- 1664
Db 17542 TCCCCCCACACACCCCGCTCATCAACCAACTCCACGACACACCCCAACCTCTCACCTAC 17601
Qy 1665 TyrProProTyr-----LeuIleArgGlyTyrProAspThrAlaAlaLeu----- 1679
Db 17602 CACCCACCCACACACCCCGCTCATCGGCACACCCCGCACCACTCTCACCCCC 17661
Qy 1680 -----GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694
Db 17662 CACTACTGGACCAACAAAGCCCGCAACCGTC-----GACTACGCCACCAACCCCAA 17715
Qy 1695 MetHisHis-AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSe 1714
Db 17716 ACCTCTCACCAACACCGGTCTCACCTACATCGAATCGGAC----- 17758
Qy 1714 rProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLe 1734
Db 17759 -CGGACACACCCCTCA----- 17773
Qy 1734 uSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAl 1754
Db 17774 -----CCACCTCTCACCAACCAACCTCTCCCCAACACCCCGCACCTCTCAC 17823
Qy 1754 aMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSe 1774
Db 17824 CTACCCACCCCGCAC-----ACCACCCCAACACCCCTCTCTCACCACTCTGCG 17874
Qy 1774 rSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSe 1794
Db 17875 AAAACACACCACTCTGGCACCCCGC-----ACTACACCCACCAACCAACCAACCC 17928
Qy 1794 rGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlySerI1 1814
Db 17929 -CACACCCACACCCCTCGAC----- 17949
Qy 1814 eLeuThrSerThrThrValGluHisAlaProIleTTP-----ArgProG1 1830
Db 17950 -----CTCCCCACCTACCCCTTCCAAACAGCAGCTACTGCTCGAAGACACAGCCGG 18005
Qy 1830 yThrGluGlnSerSerGlySer-----SerGlySerSerGlyGlyGlyGlySerSer-- 1848
Db 18006 TGCGCGATCGGTTCGGTTCGGTTCGGTTCGGGCGGAGGAGTCTCGGCGGCGGAGCGGAGA 18065
Qy 1848 ----- 1848
Db 18066 GGTGAGTCCGGTTCGGGACCGGTGGCGCCGAGGACCTCGAAGCGTTCGCGACAC 18125
Qy 1849 -----SerArgProAlaSerHi 1854
Db 18126 GCTCGCGGTCCCCCTCCCGCGCGCTGGACACGGTGGTCTCGCGCACTCTCCGCTCGCA 18185

```
Qy 885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys----- 896
Db 14116 GCACACACCTCATCCCAACGCGCCACCCCTCATGCAAAACCATGCCCCCGGCACCATG 14175
Qy 897 -GluGlyGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
Db 14176 ACCACCTTCACACACCCCGCCACCATCATCACCCACCATCTCACCGGCCCAAGAAACGAC 14235
Qy 916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnAr 936
Db 14236 CTCGCCATCGCGGCATCAACACCCCGCCACCTCCCTCGTCATCAGCGCA----- 14284
Qy 936 gLeuLeuSerProArgProSerLeuThrProThrGlyAspProArgAlaAsnAlaSe 956
Db 14285 -----CCCCCGCACCGCTCAACACATCACCA-----CCCTGCGCAACAA 14328
Qy 956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIlePro---Pr 975
Db 14329 GGCAATCAAAACCAAAACCTCCCGCCACCAACAGCGCTTCACCTCCCGCCACCAACCC 14388
Qy 975 oIleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr 994
Db 14389 ATCTCAACCAACTCCACACGACACACCAACCAACCTCACTACCAACCCACCCACCC 14448
Qy 994 oAlaProProAlaProProProGlnAsnLeuGlnProGlnSerAspAlaProGlnG1 1014
Db 14449 CTCATCCGCGCAACACCCCGCCACCACTCTCTACCCCGCCACTACTGACCAACAA 14508
Qy 1014 nProGlySerSerProArgGlyLysArgSerProAlaProProAlaAspLysGluAl 1034
Db 14509 GCCCGCAACA-----CGTCGACTACGACACCAACCAACCAACCAACCTCCACCA 14556
Qy 1034 aPheAlaAlaGluAlaGlnLys-----LeuProGlyAspPro---ProCy 1048
Db 14557 CACGCGCTCACCACTATCATCGAACTCGGACCGGACCAACACCTCACACCCCTCACCC 14616
Qy 1048 sTriThrSerGlyLeuProPheProValPro-----ProArgGluValI1 1063
Db 14617 GACAACTTCCCAACACCCCGCCACCAACCTCACCTCACCCCGCCCGCCACCAACCC 14676
Qy 1063 eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyrAlaProProGly---H 1082
Db 14677 CAAACCCACCTCTCTCACCACTCGCCAAACCAACCAACCAACCTCGGCACCCCGCCACTAC 14736
Qy 1082 iSProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrI 1102
Db 14737 ACCCACCACCA-----CAACCAACCCACCAACCAACCAACCTCGACTCGACTCCCA 14787
Qy 1102 leSerAsnPro---ProProLeuIleSer-----SerAlaLysHisProSerV 1117
Db 14788 TACCCCTTCCACACCACTACTGTGCTCCACCAACCCCGGCAAGCGGACCGCC---G 14844
Qy 1117 alLeuGluArgGln-----I 1122
Db 14845 TCACGAGCGAAGCGGTGAGCAAGCCACGACCCCATCAACCCCGCTCGGTGTCTCTC 14904
Qy 1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu-HisAla 1141
Db 14905 GTGGCAAGTCTCCGAGGAGGAGAGCAAGAGCTGTGGCGCTGTGGCGGACCACTGCG 14964
Qy 1142 LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu 1161
Db 14965 GCGCTGTGTGGCGCATGCCACTCCCGAAGTATGCTTCCGAAACAGGCGCTTCAAGAG 15024
Qy 1162 AlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro--- 1180
Db 15025 CTGGGTTTGTATCTCTCGCGCAATTCAGTTCGTATATCGACTCTTGTGCGCTGAC 15084
Qy 1181 -----ProGluSerLeu-----GlyValProThrAlaGlnGlu---AlaSerVal 1194
Db 15085 CTGCGCTTCGCGCACCGCTATCTTCGATTACCCACTCCGATCGGCGCTTTCAGTTC 15144
Qy 1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer 1214
```

```
Db 15145 CTCGCGCGCGCATCTCGGAGCGGACACAGGC-----ACGACCACTCGCTCTCGCGCTA 15198
Qy 1215 ThrArgValProSerAspSerAlaIleThr-----TyrArgGly 1227
Db 15199 ACTGCGGTTCGCCCGACGAGCGCATCGTCGGCATGCGCTCGGTATCCCGGT 15258
Qy 1228 SerIleThr----- 1230
Db 15259 GATGTACGAGCGGTTCGATGATCTCTGCGAGGTGTCAGTGTGGCCATGACGCGATCGGC 15318
Qy 1230 ----- 1230
Db 15319 GGATTCCCGACGAACCGTGGTGGACCTCGACACGCTGTACAACCGGACCGGACCCAC 15378
Qy 1231 HisGlyThr----- 1233
Db 15379 CACGGAACCAAGCTATACCCCGAGCGGGGATTCCTTTACGACGAGGCAATTTTCGATCCC 15438
Qy 1234 -----ProAlaAspValLeu----- 1238
Db 15439 GACTTCTTCGTATCAGTCCGCGTGAGGCACTGGCGATGGACCGGACGCGCTGCTG 15498
Qy 1239 -----TyrLysGly 1241
Db 15499 CTGGAACACGCGTGGGAGAGCATCGAACACGCTGCATCAACCCCGACAGCCTCCGTGGC 15558
Qy 1242 ThrIleThrArgIleIleGly-----GluAspSerProSerArg----- 1254
Db 15559 ACACCAACCGCGCTCTTCGCGCGGTGACCTACACGACTACGCGCGCGCTTTCCACACA 15618
Qy 1255 -----LeuAspArgGlyArgGluAspSerLeuProLysGly 1266
Db 15619 GCTCCGCGCAGGTTTCGAGGGGTATCTCGGCACGGAAGCGCAGCAGTATCGCTCGGT 15678
Qy 1267 HisValIleTyrGlu---GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
Db 15679 CGTGTCCCTACGCTCTCGCGCTCGAAGGTCCGCGCTCACAGTCGAC----- 15726
Qy 1286 ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAla 1305
Db 15727 ---ACTCCCTGCTCTTCGTCCTT---GGTCTCTGCA-----CCTGGCTGTGAGCGCT 15776
Qy 1306 AlaProLysArgThrTyr-----AspMetMetGluGly 1316
Db 15777 GCGTCCGCGAGGTTCATGCGCTCGCGGTGGGTGCGTCAAGTGTATGTCAACCCCGC 15836
Qy 1317 ArgValGlyArgAlaIleSerSerAla-----SerIleGluGlyLeu 1330
Db 15837 CGGTTCTGTGAGTTTTCGCGGACGCGGCGCTGCGCGTGGAGCGGCGTTCGCAAGCGGT 15896
Qy 1331 MetGly----- 1332
Db 15897 CTCGCGACGCGCTGACGCGACCGCTGGGTGAGGTGTCGGAATGTCGTGTGGAGCG 15956
Qy 1332 ----- 1332
Db 15957 GCTGTGCGACGCGCGGCTCGGTCTCGGTCAACGAATCTCTCGGTGTGCGTGTGCGGT 16016
Qy 1333 -----ArgAlaIlePro----- 1336
Db 16017 CAATCAGACCGTTCGAGCAACCGGCTGACGCGCGCCCAACGCGCGCTCCAGGAGCGGT 16076
Qy 1337 -----ProGluArgHis-----Ser 1341
Db 16077 CATCCGCTGGCTGCGCAACCGGACCTGACCCCGCGCAGCTCGATCGGTGGAGGC 16136
Qy 1342 ProHisLysLeuLysGluGlnHisHisIle---ArgGlySerIleThrGlnGlyIlePro 1360
Db 16137 CCACGGCAC-----CGGCACCACTTTGGGCGACCCGATCGAGGCCCGGCGCTCT 16187
Qy 1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArg 1380
```

Db 12012 GCCTGCTACCTGCGGACCTGACCGCACTCCGCTGCTGTACGAGTGTGTCAGCGTGA 12071
Qy 350 tGlnSerArgValGlyGlnArgGly-----SerGlyLeuSerMetSerAl 365
Db 12072 GCAGGAGCGGATCGGATGTGGGATGGCGTGTCTGTTACCCAGCGGGCGAGCTCACC 12131
Qy 365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu----- 383
Db 12132 CACGCGAGTGTGCGATCTGCTCAAGTCCACAGCGAGCGTATCGGGAGTTCGCCGCCGA 12191
Qy 384 -----AsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTy 400
Db 12192 CCTGTGATGAACTGAGCG-----CTCTA 12218
Qy 400 rAspAlaAspGlnGlnArg-----IleTysPheIleAsnMetAsnGlyLeuMetAlaAs 418
Db 12219 CGACCCGACCCCGCTCAGAACCCAGTTACACGCGCAGCGAGGGTTTCTCTATGA 12278
Qy 418 p-----ProMetLysValTyrLysAspArgGlnValMetAsnMetTr 432
Db 12279 CGCGGGCGACTTCGACGCGCGTCTTCGAGTTGTCAACCGCTGAGCGCTGGCAATGGA 12338
Qy 432 pSerGluGlnGlu-----LysGluThrPheArgGluLysPheMetGln----- 446
Db 12339 CCGCAGCAGCGCTGCTGCTGAACCACTTGGGAAACGTTGAAACAGCGCGGATCGA 12398
Qy 447 -----HisProLysAsnPh 451
Db 12399 CCGAGGTCCATCGCGGAAGCGGACCGGGTTTCGTGGGNTCAATCGGAGGACTA 12458
Qy 451 e-----GlyLeuIleAlaSerPheLeuGluArg-----Ly 461
Db 12459 CACCACCGGATACAC-ACATCAGCCCTCAAAACGAGTCAGGCGTACCTCTCACTGGCA 12517
Qy 461 sThrValAlaGluCysValLeuTyrTyrTyrLeu-----ThrLysLysAsnGl 477
Db 12518 GCGCGGCAAGCATTGGCTGAGCGGTATCTCTACAACTTCGGGCTCGAAGGCGCTCGCA 12577
Qy 477 uAntTyrLysSerLeuVal----- 483
Db 12578 TCATATCGACACCGGTGTCTCTCGCTGCTGCGCCCTGTCATCGGCTGCGCAAGCGC 12637
Qy 484 -----ArgArgSerTyrArgArgArgGlyLysSerGl 494
Db 12638 TCGGTCCGGTGAATGACCATGCGCTCGCAGCGCGCTCTCGTATGCGCACTCCCT 12697
Qy 494 nGln----- 495
Db 12698 TCGTCTTACCGAGTCTCTCGCCAGCGGGGCTGCGCCGACAGCGCGGTGCAAGCGGT 12757
Qy 495 ----- 495
Db 12758 TTTCGGGGCGGCGGACCGGCTGGTCCGAGGGTGTGGGATGCTGCTGTTGGAGC 12817
Qy 496 -----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Db 12818 GGCTCTCCGACGCGCGCAAGCGTCAACGCTGTCTGCGCGTCTGCGCGGAGCGCGC 12877
Qy 509 nGlnProMetProArgSerSerGlnGluLysAspGluLysGluLysGluLysGluLys 529
Db 12878 TCACCCAGGAGCGGCGAAGCAAGCGCTGACCGCAACCCCAACCGGTCTTCAAGTCAAGG 12937
Qy 529 aGluLysGluGluLysPro-GluValGluAsn-----AspLysG 543
Db 12938 T-----CATCGCGAGGCTTTGGCCCAACGACACCTCTCCCTCGCGGATGTCG 12985
Qy 543 luAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluA 563
Db 12986 ATCGGTGGAGGCGCCACGCGACCGGACCACTCCCTGGGCGACCGGATCGAGGCTCAAGCCC 13045
Qy 563 laValAlaSerLysGlyArgLysThrAlaAsnSerGln-----GlyArgArgL 579
Db 13046 TCGTGAAGCGCTACGGTCAAGACCGGCCCAACGCGCGCCCTCTGTGGCTCGGAACCCCTCA 13105

Qy 579 yGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer----- 592
Db 13106 AGTCCAACTCGGCACTCCATGGCGCTGCGGGTGTGGCGGGTTCATCAAGATGGTGA 13165
Qy 593 -----GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuA 610
Db 13166 TGGCGTGGCGAATGTCTGCTGCCCGG-----ACGTTGCATGTGG 13207
Qy 610 snGluSerSer-----ArgTyrThrGluGluMetGluThrAlaLysLysGlyL 627
Db 13208 ATGAGCGCTGCGCGCATGTGGACTGTCGCGGGTGGTGCAGCTG-----C 13255
Qy 627 euLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal- 646
Db 13256 TGACGGAGAC-GGT-----GCCGTGCGCGGGGAGGGCGGTACGCGCG 13302
Qy 647 -----SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluI 665
Db 13303 GCAGGAGTGTCTATTCGCGGTCAGCGGCACCAACGCCACCTCATCTCTCGAGGAAGCA 13362
Qy 665 leLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysL 685
Db 13363 CCGCGCCCAACAATCCCGTCAGACACACCGCGCGAGCGTCCCGGGAGAAATCAGCGGCC 13422
Qy 685 ys-AlaProAlaAlaAlaSerGluGluAlaAla-----PheProPro 698
Db 13423 GACGAGGATGCGGTAGTGGCGATGAGGTCTGTCGCGGAGTCCAGGGGTGTGCGCGTGG 13482
Qy 699 valValGluAspGluMetGluAlaSerGlyValSerGlyAsnGluGluMetVal 718
Db 13483 CTGCTGCGGCCAAGTCGCGCGCGC-----CTGCGC 13515
Qy 719 GluGluAlaGluAlaLeuHisAla-SerGlyAsnGluValPro-----ArgGlyGluCy 736
Db 13515 GCCCAGGCCAGGCCCTTGACGCCACCTCACGACACCCCGCTCGACCTCGCGAC 13575
Qy 736 sSerGly-----ProAlaThrValAsnAsnSerSerAspThrGluSerIl 751
Db 13576 GTCGGTATACCTCGCCCAACGCGCGCGCTTTCGACACCGCGCCACCTCATCGCC 13635
Qy 751 ePro-----SerProHisThrGluAlaLysAspThrGlyGlnAsnGlyProLysPr 769
Db 13636 GCGCAGCGGACACCTCTCTGCAAGCACTCCAGGCACTCGCCGCGAGTCCCGCGAGAACCCACCC 13695
Qy 769 oProAlaThrLeuGlyAlaAsp-----GlyProProGlyProPro----- 783
Db 13696 GCGTCTATCCACAGACGCCCCCAGCGGGGACCGGGACCGGGAGCGCGGAGAAAGACC 13755
Qy 784 -----ThrProProArgArgThrSerArgAlaProIleGluProThr----- 797
Db 13756 GCATTCTGCTCGGACAGGCGACCAACGCGCGGATGCGCCACGCGCTTACCAC 13815
Qy 798 -----ProAlaSerGluAlaThrGlyAlaPro----- 806
Db 13816 ACCACCCCGTCTTCGCGCGCGCACTCAACAGACATCTGCACCCACCTCGACCCCCACCTC 13875
Qy 807 -ThrProProAlaProProSerProSerAlaProProProValProValProLysGluGl 826
Db 13876 GACCACCCCTCTCTCCCTCTCTCACCAGGACCCCAACACCCAGGACACACCCCTC 13935
Qy 826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGluGluGlnLysProPr 846
Db 13936 GAAGAAGCGCGCACTGCTCCAGCAGACCGCTAGCGCCCGCGCCCTCTTCGCGCTTC 13995
Qy 846 oAlaAlaGluGluLeuAlaValAsp-----ThrGlyLysAlaGluProValLysSerGl 865
Db 13996 CAGGTCCGCTCCACCGCTCTCTCACCGAGCGGTACACATCACCCCTACTACTAGCC 14055
Qy 865 uCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaAlaGluAla 885
Db 14056 GGACACTCTCTCGCGCAATATCACCAGCGGCCACCTCGCGCGCATCTCACCTCAGCGAC 14115

| | | | |
|---|--------|--|-------|
| Percent Similarity: | 30.56% | Conservative: | 308 |
| Best Local Similarity: | 21.31% | Mismatches: | 1176 |
| Query Match: | 5.17% | Indels: | 1152 |
| DB: | 3 | Gaps: | 152 |
| US-09-522-753-5 (1-2517) x AAA92301 (1-30690) | | | |
| Qy | 44 | HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArg | 63 |
| Db | 11193 | CATCATCCGTTCCGTCATCCATCCATCGACCCCGGA | 11234 |
| Qy | 64 | ArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHis | 83 |
| Db | 11235 | ACTGCCCTCTTCGATGCGC | 11285 |
| Qy | 84 | LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPheI | 103 |
| Db | 11286 | CGCCGACCTGCTCCGCCCGCCCTGCGCCCTG | 11336 |
| Qy | 103 | eGluSerLysArgProArgLeuLeuLeu-ProAspProLeuLeuArg-ProSerP | 122 |
| Db | 11337 | GCCCGCCACCCGCGCCGACACCCGACACCTACCGTGTGCGGACACACGCGC | 11396 |
| Qy | 122 | roLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspSerLeu | 142 |
| Db | 11397 | CCAGCTGCACGCCCGCTGGCGCGCCAGACACAGAACACAGCACACCCTCTCTGC | 11456 |
| Qy | 142 | hrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeuGlu | 162 |
| Db | 11457 | CTGTGTCC---GCTCCACATCGCCACGCTCTG-GGCCACACACACCC | 11506 |
| Qy | 162 | euValProArgLeuSerLysGluLeuLeuGlnAsnMetAspArgValAspArgG | 182 |
| Db | 11507 | CCATCCCGCC | 11533 |
| Qy | 182 | luIleThrMetValGluGlnGlnIleSerLysLeuLysGlnGlnGlnLeu-Glu | 201 |
| Db | 11534 | ACCTCGCTTCGACTCCCTCACCGCGCTCGAACTACGCAACCGGCTCTCCGCCACCCG | 11593 |
| Qy | 202 | GluGluAlaAlaLysProProGluLysProValSerProProIle | 219 |
| Db | 11594 | GACTCGCTCCCGCCACCTCGCTTCGACACCCCAACCCACACCTCCACCCAC | 11653 |
| Qy | 220 | GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg-LysLysAlaG | 239 |
| Db | 11654 | ACCTCCACACACAACTCC | 11693 |
| Qy | 239 | uAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu-ProLeuTyrAs | 258 |
| Db | 11694 | CGCCCGCTGTGGCGAGCTCGACAACTCGAATCGCCCTCTC | 11738 |
| Qy | 258 | nGlnProSerAspThrArgGln | 267 |
| Db | 11739 | CGCCCTCGACAAACCGACAGCGCCAGCGAAGAGTCACTCGCGGTGAAGTCACCTCA | 11797 |
| Qy | 267 | sGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgAr | 287 |
| Db | 11798 | TGTTAGGTGGAGC-CACCCAGCATCGACGCGGAAAGCGCTGATCAGCAGCAGAG | 11856 |
| Qy | 287 | gAsnHisAlaArgLysGlnTrpLysGlnLysPhe | 298 |
| Db | 11857 | TTACAT-----CGGCACAGAGGTGAGATTTTCAATTCATTGACAAACGACCTCGC | 11910 |
| Qy | 299 | -CysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs | 318 |
| Db | 11911 | CTGTCTGAACCGAGCGCTGCCACCTCCGC | 11952 |
| Qy | 318 | naSnProArgArgArg-----AlaLysGluSerLysValArgGluTyrG | 334 |
| Db | 11953 | GGCCCTGTAGGACGTGAATGGCAATGGCAATGAGCGAGGCTCTCTGGAATACCTCAA | 12011 |
| Qy | 334 | uLysGlnPheProGluIleArgLysGlnArg-----GluLeuGlnGluArgMe | 350 |

```
QY 1785 -----LeuThrLysProThr-----ThrThrSerSerSerGlu 1795
Db 40031 CGCGGTGCGCAACCGCGCAACCGCGCGCGCGCGTGGACACAGCAGCGCGCGGT 39972
QY 1796 ArgGluArgAspArgAspArgGluArgAsp-----ArgAspArgGluArgGluLys 1812
Db 39971 TGCACCGCGCGCGCGTCCCAACCGAGGAAACAAATCCGCGCGCGCGCGCGCGCG 39912
QY 1813 SerIleLeuThrSerThrThrThrValGlu-----HisAlaProIleTrpArgProGly 1830
Db 39911 CGCTGCGCAACACAGCCCTGCGCTACCGCGCGCGCGCGCGTCCCAACCGGTGTCCAGGT 39852
QY 1831 ThrGluGln-SerSerGlySerSerGlySerSer-----GlyGlyGlyGlyGlySerSe 1848
Db 39851 TTTTCGCGGTACCGCGCGGTCCCGCGCGCGCGCGCGCGCGAGCATTCGCGCATTC 39792
QY 1848 rSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGl 1868
Db 39791 CGCGCGCGCGCGCAACCGCGCGTCCGCTAGTGTGAGCGCGCGGTCCCGCGGTGAC 39732
QY 1868 naSpAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db 39731 CTCGCGCGGACCGATCAGCCCGCGC---CAGCAGTCCGCGGTCCCGCGCGC----- 39683
QY 1888 rAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSer-ProValA 1908
Db 39682 -----CCACCGCGCGCGCGTGC-----GCCACCAACAGCGCGGTTC 39645
QY 1908 rgProAlaAlaThrProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyV 1928
Db 39644 CACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39607
QY 1928 alTyProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgP 1948
Db 39606 -----CGACCGCGCTCCCGCGCGGTGCGCGCGAGT-----ACGTGCTGCTC 39561
QY 1948 roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
Db 39560 CGCGGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39507
QY 1968 lyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProV 1988
Db 39506 CC---CCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39450
QY 1988 alSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAlaAs 2008
Db 39449 TACCC---CACAGCCCG---GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39396
QY 2008 erProAspProAlaProAlaProAlaSerAlaSerAspProHisArgGluLysThrGlnS 2028
Db 39395 CGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39355
QY 2028 erLysProPheSerIleGlnGluLeuLysArgSerLeuGlyTyrHisGlySerSerT 2048
Db 39355 ----- 39355
QY 2048 yrSerProGluGlyValGluProValSerProValSerProSerLeuThrHisAspL 2068
Db 39354 -----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39321
QY 2068 ysGlyLeuProLysHisLeu-LeuGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
Db 39320 CGGGGTGCGCG-----TTGCCATCAGCGGTGCCCGGTTCAGTGTGCGGTGCGCGGT 39267
QY 2088 ProLysGlnProGlyProValLysLeuGlyGlyGlyAlaAlaHisLeuProHisLeuArg 2107
Db 39266 TGATCGCGCGCGCGCGCGCGTGT---GTTTACGGGTGTGCACGCGAGAG-GCGGTGCGCGG 39211
QY 2108 ProLeuProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLys 2127
Db 39210 GCCTTGAAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG----- 39160
```

```
QY 2128 GlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyr 2147
Db 39159 -----GTATCACCGCGGTGCGTCCGCTT----- 39136
QY 2148 ThrArgHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhePro 2167
Db 39135 -----CCCGCGATGCGCGGTGCGCGCGGTTCGCGCGTTCGCGATCAGCACCGCGTCCGCGCG 39085
QY 2168 GlyAlaSerCysProValLeuAspLeuArgArgProPro----- 2180
Db 39084 GCCCGCGCGGTACCGCGGTACCGCGCGCTTGCCACCGCGCGCGCGCGCGCGCGGTTA 39025
QY 2181 -----SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPro 2198
Db 39024 CGGATCAGCAACCGCGCGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTCGCGCG 38965
QY 2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal-----Leu 2215
Db 38964 AACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38906
QY 2216 GlyGlyGlyGluAspGlyIle---GluProValSerProGluGlyMetThrGluPro 2234
Db 38905 GCCCGTCTCCCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38846
QY 2235 GlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluPro 2254
Db 38845 CAGCACACCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38810
QY 2255 Ser-----ArgMetGlySerLysSerProGlyAsnThrSerGlnPro 2268
Db 38809 ACCGTACACCTAGCGCGCGTCCCGCGGTATCCACCGGTGCCCGCGCGCGCGCGCG 38750
QY 2269 ProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGln 2288
Db 38749 CCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38717
QY 2289 GluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGln 2308
Db 38716 ---GTCGTTGAGAAAGCACGTCGCGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCG 38660
QY 2309 ProGly-----ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeu--- 2324
Db 38659 CCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38600
QY 2325 -----MetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
Db 38599 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38546
QY 2342 LeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerPro 2361
Db 38546 ----- 38546
QY 2362 ProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro-----Ala 2379
Db 38545 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38486
QY 2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu----- 2394
Db 38485 CAACAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38426
QY 2395 -----ThrSerProGlyGly-----GlyGlyLysAlaLysValSer 2406
Db 38425 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38366
QY 2407 -GlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspAr 2426
Db 38365 GGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38309
QY 2426 gProProSerValSerSerValHisSerGluGlyArgCysAsnArgThrProLeuTh 2446
Db 38308 CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38264
QY 2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe---ProTyrAs 2465
```

| | | | | | |
|----|-------|--|-------|---|-------|
| Db | 41816 | CGTTACCATATTGATCAGCTCTGCTGCGAGGTGTGACAGGGCGGAGGTGCTGGCGGGG- | 41758 | CGCGTTCCTGATCAGCGGGCGCGGCTGGCGCCACGAAATTGCGCTTGATCGGGGCCA | 40922 |
| Qy | 1129 | etSerValGlnLeuHisValProTyrSerClnHisAlaLysAlaProValGlyProValT | 1149 | --LysTyrAspThrGlyAlaSerThrThrGlySer-LysLysHisAspValArgSerLeu | 1476 |
| Db | 41757 | --GGCGTAAAGCGGTTCAGCGCCCAACAGCACCCCGCGGTACCGCGATGCGCGCTTGC | 41700 | GCAGCGCGCACAGCGCGCGGCTCGCGCGCGGATACGGCGCTCCCGCGCGCTCAGGG | 40862 |
| Qy | 1149 | hrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysG | 1169 | lleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAla | 1496 |
| Db | 41699 | CAAGGTGCGCGCGGTCCCGCCGCGCGG-----CGTTGCGCGCGTTCGCGA | 41652 | CCTGCAGCAACTGGCGGTAAACCGCGCG-----CCTGTG----- | 40826 |
| Qy | 1169 | lnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThra | 1189 | ArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly----- | 1513 |
| Db | 41651 | TCACACGCGGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 41604 | CGCTCGCGCGCTGTAGCGCTTGGCGCTTGGGCAC | 40793 |
| Qy | 1189 | la---GlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerI | 1208 | ThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGlu | 1532 |
| Db | 41603 | CGGTGCTCGCGCGCGGTTCGCGCGCGGTTCGCGCGGTTCGCGATCAACCGCGGGCGCG | 41544 | CGAACACACCGCGCAATCTGTGACACCTCATGTGCGACCCCGCGGTCAACAATCCGTCG | 40733 |
| Qy | 1208 | leThrLys-----GlyIleProSerThrArgValProSerAspSerAlaIleThrT | 1225 | LeuGlyLysProArgLysSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGly | 1552 |
| Db | 41543 | CGCGCGCGCGCGCGCGGTCTCCGACGCTTCTACTAATGCGCGCGCGCGCGCGAG | 41484 | TGTTACGCGCG----- | 40715 |
| Qy | 1225 | yrArgGlySerIleThrHisGlyThrProAlaSerValLeuTyrLysGlyThrIleThra | 1245 | HisLeuPro-----ArgGlySerProValThrMetArgGluProThrPro | 1567 |
| Db | 41483 | CGCGCGGAGCGAAGAACATCGCGCGGTTCGCG----- | 41452 | CATTAGCGGAGCTAATCGCGACCGCATCCCGCCCAATCCGCTCGCGCGCACCA | 40655 |
| Qy | 1245 | rgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProL | 1265 | ArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSer | 1587 |
| Db | 41451 | -----GCCGGTTCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 41397 | GTCG | 40601 |
| Qy | 1265 | ysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetS | 1285 | ThrProArgGluIleAla----- | 1593 |
| Db | 41396 | CGCGCGCGCGCATCACAGTGGCGAGCGTCCGCGGTTCGCGCG----- | 41353 | TGTTGCGCGACTACACGACGACACCCGCTACCGGTAGGAAATTAATTGCAGAAATCAACGA | 40541 |
| Qy | 1285 | erValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrA | 1305 | LysSerProHisSerThrValProGluHisHisProHisProHisSerPro | 1610 |
| Db | 41352 | -----GCCCG | 41325 | TATCGCGATCCCGCACCGCATCGCGCTTCACACGTTTCGCG----- | 40490 |
| Qy | 1305 | laAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerA | 1325 | TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro--- | 1629 |
| Db | 41324 | CCCGACCC----- | 41310 | GCTGCTGGACGGCTCAGCGGTGCGCGCGGTTCGCGACAAACGGCGCGGTGCGCGGC | 40430 |
| Qy | 1325 | laSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisL | 1345 | LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaA | 1647 |
| Db | 41309 | CGCGCTTCCTCCCG | 41257 | ACAGGCG | 40376 |
| Qy | 1345 | euLysGluGlnHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValG | 1365 | AlaTyrTyrLeuProArgHisLeuAlaProAspProThrTyrProHisLeuTyr-ProPr | 1667 |
| Db | 41256 | CCCGCGGTGGCACCGC-----CGTTTCGAGATCC | 41227 | GCG-----ACGGGCTCTGACAAAGCGCTCGAGCGCGCGCGGATCTTCGCGGAG | 40328 |
| Qy | 1365 | luAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArgGluGlyThrProP | 1385 | oTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleLea | 1687 |
| Db | 41226 | CGCGACCG | 41168 | ATACCTGGCG | 40292 |
| Qy | 1385 | ro---ProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeu--- | 1402 | naSpTyrIleThrSerGlnGlnMethHisHisAsnThrAlaThrAlaMetAlaGlnArgAl | 1707 |
| Db | 41167 | CG | 41108 | CG | 40242 |
| Qy | 1403 | -----GlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrV | 1418 | aAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaG | 1727 |
| Db | 41107 | ACAGCG | 41070 | CCGTGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 40182 |
| Qy | 1418 | alLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProG | 1438 | y-Pro-----ArgGlyIleLeuAspLeuSerGlnValProHisLeuProValLeuValP | 1745 |
| Db | 41069 | -----CCG | 41018 | CGCGCGCGCGTTCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 40125 |
| Qy | 1438 | luLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeu- | 1457 | roProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlap | 1765 |
| Db | 41017 | GCGCGCGGTGGCTCCGCTCCG-----GGGGCGCGCGTGG | 40982 | CGCGGTGCCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 40086 |
| Qy | 1457 | ----- | 1457 | roGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThr-His | 1784 |
| Db | | | | CGGCGCGCG-----CCGAAGCCCGACGCGCGCGCATTTGCGCGCGCGCGCGCGCG | 40032 |

| | | | |
|---|-------|--|-------|
| Query Match: | 5.1% | Indels: | 652 |
| DB: | 4 | Gaps: | 95 |
| US-09-522-753-5 (1-2517) x AA199683_08 (1-110000) | | | |
| QY | 546 | LeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla | 565 |
| DB | 43605 | GTTCGTCGAAGGTATGGCGCTGGCGGTGACGGTTCGACGAATGCGAATGCTCTGC | 43546 |
| QY | 566 | SerLysGlyArgLysThrAlaAsnSer | 578 |
| DB | 43545 | TCCGCGTGAAGCAATTCGAAGATGCGTGCATTCGTTGTCAGCGCGGAGAACACGCCCA | 43486 |
| QY | 579 | -----LysGlyArgLysThrAlaAsnSer | 592 |
| DB | 43485 | TACCAGGTTCCGACACTGTGAGGGGTGCTGACGCGGAGCAGTCGATCGACACGAGGGCG | 43426 |
| QY | 593 | Glu-----GluAlaIleThrProGln | 600 |
| DB | 43425 | GACGGTTCCGATGATTCGGGTGTCACCGCGGTGTCAGCGCAGCAGCAGTCGTCGAG | 43366 |
| QY | 601 | -----SerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTyr | 615 |
| DB | 43365 | AAGGAACATTCGCTTCGCGCTATTTGGGGTTACTTCGGGAGTTCAGTTCAGGCC | 43306 |
| QY | 616 | ThrGluGluMetGluThrAla | 623 |
| DB | 43305 | TCGTGACGACATTCGGAACAAGCTCTATTTCAGGTTCAGTTCGCTCGCGCGGACCG | 43246 |
| QY | 624 | -----LysLysGlyLeuLeuGluHisGlyArg | 636 |
| DB | 43245 | GTCCGCGCGTCGAACACCGGACGGTC-----GGCGGCGCGCGGACGTGGCGGCG | 43192 |
| QY | 637 | Ile----- | 637 |
| DB | 43191 | GCGAATCCCGCGAAGGGCGTCTCGATCAGCGCCCGCAGCGATTGACACGCTCGCGG | 43132 |
| QY | 638 | AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys | 657 |
| DB | 43131 | GCCCGCGTTTGGCGGACGAGTATTTCA----- | 43102 |
| QY | 658 | LysArgGlnAsnLeuAspGluIleLeu-----GlnGlnHis | 669 |
| DB | 43101 | -----TCGAGATTGACACCGGTGGTGGCGATGATGCACACGGTAACGATCAACAGCAT | 43048 |
| QY | 670 | LysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAla | 689 |
| DB | 43047 | CATCTCGGTC--GAGCGCGCGAAGCGATGCTGTCAACGCGCGCGGTGCGAGCGCCG | 42989 |
| QY | 690 | AlaSerGluGluAlaAlaPheProValValGluAspGluGluMetGluAlaSerGly | 709 |
| DB | 42988 | CGGCTACATCAACCGTTTCAGCCCGTCTCCAGCAGCAGCGCGCGCGCGCGCGG | 42929 |
| QY | 710 | ValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsn | 729 |
| DB | 42928 | GCCGCTGGG----- | 42920 |
| QY | 730 | GluValProArgGlyGluCysSer-----GlyProAlaThrValAsnAsnSer | 745 |
| DB | 42919 | -----GCCGGGTGACCCCGGATTTCCCGGGCTTACCCTGTTCACCGCGTT | 42872 |
| QY | 746 | SerAspThrGluSerIleProSerProHisThrGluAlaAla----- | 759 |
| DB | 42871 | GCCGCGTTCCCGCGTTCGCGATCAGCAGCGGCTGCGCGCGCGCGCGCTTCGCGCC | 42812 |
| QY | 760 | -----LysAspThrGlnAsnGly-----ProLysProProAlaThrLeuGlyAl | 775 |
| DB | 42811 | GATGTGAAAGACCGGACAAACCGGTGGCGCGCGCGCGCGCGCTCACCTCCGAGACC | 42752 |
| QY | 775 | AspGlyPro-----ProGlyProProThrProPro-----ArgArgThrSerAr | 791 |
| DB | 42751 | GAACGCTCCGACTTCGCGCGCGCGCGCGCTCCCGCGGTGGCGCGCGCGCGCGG | 42692 |
| QY | 791 | gAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProAl | 811 |
| DB | 42691 | GGCG | 42635 |
| QY | 811 | aProProSerProSerAlaPro-----ProProValValProLysGluGly | 827 |
| DB | 42634 | CCCG | 42575 |
| QY | 827 | sGluGluGluThrAlaAlaProProValGluGluGluGluGluGluGluGluGluGlu | 847 |
| DB | 42574 | GCGGAAGACAGCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 42539 |
| QY | 847 | aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCys | 867 |
| DB | 42538 | GACCAACG | 42506 |
| QY | 867 | rGluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu | 887 |
| DB | 42505 | -----GCTGCG | 42453 |
| QY | 887 | aGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAla | 907 |
| DB | 42452 | CG | 42393 |
| QY | 907 | sSerSerGlyAlaProGln-----AspSerAspSe | 917 |
| DB | 42392 | TGCG | 42333 |
| QY | 917 | rSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLe | 937 |
| DB | 42332 | CGTTCACCTGCG | 42305 |
| QY | 937 | uLeuSerProArgProSerLeuLeuThrProThrGlyAsp-ProArgAlaAsnAlaSerP | 957 |
| DB | 42304 | -----AGCCCG | 42270 |
| QY | 957 | roGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaIleProIleG | 977 |
| DB | 42269 | CGGCTCCCGCGTTG-----GCGAACAGCGCGCGCGCGCGCGCGCGCGCGCG | 42238 |
| QY | 977 | lnValThrLysValHisGluProProArgGluAspAlaProThrLysProAlaProp | 997 |
| DB | 42237 | -----CG | 42198 |
| QY | 997 | roAlaProProProGlnAsnLeuProGluSerAspAlaProGlnGlnProGlyS | 1017 |
| DB | 42197 | CG | 42168 |
| QY | 1017 | erSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAla | 1037 |
| DB | 42167 | CGGACCG-----GCGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 42117 |
| QY | 1037 | laGluAlaGlnLysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProV | 1057 |
| DB | 42116 | CACCGCTCCG | 42057 |
| QY | 1057 | alProProArgGluValLysAlaSerProHisAlaProAspProSerAlaPheSer | 1077 |
| DB | 42056 | TCCCGCGAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 41997 |
| QY | 1077 | yrAlaProProGlyHisProLeuProLeuGly-----LeuHisAspThrAlaArgProV | 1095 |
| DB | 41996 | AGATCCCG | 41937 |
| QY | 1095 | alleuProArgProProThrIleSerAsnProProLeuLysSerSerAlaLysHisP | 1115 |
| DB | 41936 | GCCCGCGGTTCGCGCGGTTCGCGAACACAGCGCGCGCGCGCGCGCGCGCGCGCG | 41877 |
| QY | 1115 | roSerValLeuGluArgGlnIle----- | 1122 |
| DB | 41876 | CGGAGTGGCGTGGCG | 41817 |
| QY | 1123 | -----GlyAlaIleSerGlnGlyM | 1129 |

| | | | | | | | |
|---|---|--|-------|--|--|---------------------------------|-------|
| D | b | | 19112 | TCCCCCACAACACCCCTTACCAC | | -----CGTCATCCACACCGGCA | 19156 |
| Q | y | | 2150 | s-----HisProGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPh | | -----CGTCATCCACACCGGCA | 2166 |
| D | b | | 19157 | TCTCGACGAGCCACTTCACCAA | | -----CCTCACGCCCACTCAACAACG | 19207 |
| Q | y | | 2166 | eProGlyAlaSerCySPtoValLeuAspleuArgProProSerAspLeuTyLeuPr | | -----CCTCACGCCCACTCAACAACG | 2186 |
| D | b | | 19208 | TCTTCGGCGCCAAGGCCA | | -----CAGCGCCCACT | 19240 |
| Q | y | | 2186 | oProProAspHisGlyAla | | -----ProAlaArgGlySerPro | 2198 |
| D | b | | 19241 | TCCACCACTACCCCAACACACCCCCTCAAGCCCTTGCTCTACTCTCTCCGCGCGG | | -----CCTCACGCCCACTCAACAACG | 19300 |
| Q | y | | 2199 | -----HisSerGluGlyGlyLysArgSerPr | | -----HisSerGluGlyGlyLysArgSerPr | 2207 |
| D | b | | 19301 | CCACCTTCGGCGCACCGGCCAACCACTACGCGCGCAACGCCTTACCTCGACGCC | | -----CCTCACGCCCACTCAACAACG | 19360 |
| Q | y | | 2207 | oGluProAsnLysThrSerValLeuGlyGlyGluaspGlyIleGluProValSerPr | | -----CCTCACGCCCACTCAACAACG | 2227 |
| D | b | | 19361 | TGCCCCA | | -----CAGCGCCCACT | 19369 |
| Q | y | | 2227 | oProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyProLeuLeuTyAr | | -----CATCG | 19405 |
| D | b | | 19370 | ACGCCACACCCACCACTCCCGCCACCA | | -----CATCG | 19405 |
| Q | y | | 2247 | gAspGlyGluInThrGluProSerArgMetGlySerLysserProGlyAsnThrSerGl | | -----CATCG | 2267 |
| D | b | | 19406 | CCTGGCGCACTGGCAAG--AAACGGACTGGCGACTGGTCAAGTCAGCAACATCTCC | | -----CATCG | 19462 |
| Q | y | | 2267 | nProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLY | | -----CATCG | 2287 |
| D | b | | 19463 | GCGCGCGCGGA-TGTTCGCATTCGCCCGCGAGTTGGCGGTCAAGCTTGTGACGGCGG | | -----CATCG | 19521 |
| Q | y | | 2287 | sGlnGluIleAsnLysLeuAsnThrHis | | -----AsnArgAsnGluProGl | 2303 |
| D | b | | 19522 | ATCGCGCGCGCGCGCGAGTCTCTCGTCGCGGATACGCTGGAGAATAATTGGACCG | | -----AsnArgAsnGluProGl | 19581 |
| Q | y | | 2303 | uTyAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlalleThrGlyThr-- | | -----AsnArgAsnGluProGl | 2322 |
| D | b | | 19582 | GTTCTCTCC-AGCAAGTCGTGGTCTTGTCTGAGGACCTTCCCGAGGACAGGAACTGA | | -----AsnArgAsnGluProGl | 19640 |
| Q | y | | 2323 | -----GlyLeuMetThrTy | | -----GlyLeuMetThrTy | 2327 |
| D | b | | 19641 | GGAGGCGCGGTFACGTTGAGCAGACGAGAGACCAACCTCCGGCAACTCTCATGGG | | -----GlyLeuMetThrTy | 19700 |
| Q | y | | 2327 | rArgSerGlnAlaValcInGlnHisAlaSerThrAsnMetGlyLeuGluAlaIleAr | | -----GlyLeuMetThrTy | 2347 |
| D | b | | 19701 | TCGGTCAGTTCGAGCAGGAGAGAGCTG---- | | -----CTCAGCTCGTCCGCACTCCATCTC | 19754 |
| Q | y | | 2347 | glyAlaLeuMetGlyLysTyaspGlnTrpGluGluSerProProLeuSerAlaAsnAl | | -----CTCAGCTCGTCCGCACTCCATCTC | 2367 |
| D | b | | 19755 | CGCGGAGTAGTCTCGGCGC--GACCACTCCGAGGCATCCCGCGCGG--TCGGCTGTTC | | -----CTCAGCTCGTCCGCACTCCATCTC | 19810 |
| Q | y | | 2367 | aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs | | -----CTCAGCTCGTCCGCACTCCATCTC | 2387 |
| D | b | | 19811 | GGGATCTAGGTTGCGACTCTTCGCGCGGTGGAGCTTCGACCAACCTCCGAGCAGACA | | -----CTCAGCTCGTCCGCACTCCATCTC | 19870 |
| Q | y | | 2387 | pGly-----ArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLy | | -----CTCAGCTCGTCCGCACTCCATCTC | 2404 |
| D | b | | 19871 | CGGAGCTGGCTCTCGGACGACTCTGCTTCGATTACCCAGCCCCACCAAGCTCGGCC | | -----CTCAGCTCGTCCGCACTCCATCTC | 19930 |
| Q | y | | 2404 | sValSerGlyArgProSerSerArgLysAlaLysSerProAla | | -----CTCAGCTCGTCCGCACTCCATCTC | 2418 |
| D | b | | 19931 | AATTTCTGCTCT-CCGAGATTCGGG--AGTTCACGCCGACAACTCAACTCCGCTTCG | | -----CTCAGCTCGTCCGCACTCCATCTC | 19986 |
| Q | y | | 2419 | ----ProGlyLeuAlaserGlyAspArgProProSerValSer--SerValHisSerGl | | -----CTCAGCTCGTCCGCACTCCATCTC | 2436 |
| D | b | | 19987 | CGACCCCGGACAGCTCGATGAGCGATCGCCATCGTTGGCATGCGCTGTGCTTCC | | -----CTCAGCTCGTCCGCACTCCATCTC | 20046 |
| Q | y | | 2436 | uGlyAspCyAsnArgThr | | -----ProLeuThrAsnArgVa | 2449 |

| Db | 20047 | GGCGGAGTCACTCGCGGACGACTTCTGGGATCTGATCTCTCCGAGGAGGACCGCATC | 20107 |
|---|---------------|---|---------|
| Qy | 2449 | lTtpGluAspArgProSerSerAlaGly | 2464 |
| Db | 20107 | GGCGGATTCCTCCACGACCGCGGCTGGGACCTGGACACGCTCTACGACCCCGACCCGAC | 20166 |
| Qy | 2464 | rAsnProLeu | 2468 |
| Db | 20167 | CACCCCGGACCTGTACACCGAAAGCGGGATTCCTCTACGAGCGAGGCCACTTCGAC | 20226 |
| Qy | 2468 | eMetArgLeuGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAl | 2486 |
| Db | 20227 | GCCGAATTCCTTCGSCATCAGCCCCCGCGAAGCCCTCGCATGGACCCCGGACCAACGACTC | 20286 |
| Qy | 2486 | aglySarglyProLeuAlaGlyProHisAlaTrpaspGluGluProLysPro | 2504 |
| Db | 20287 | CTCCTCGAACCCGCTGGGAAACCATCGAACACGCGGGCATCAACCCCGACACCC | 20341 |
| RESULT 30 | | | |
| AAI99683_08/c | | | |
| Continuation (9 of 44) of AAI99683 from base 800001 (Mycobacterium tuberculosis | | | |
| WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683 | | | |
| WP | Fragment Name | Begin | End |
| WP | AAI99683_00 | 1 | 110000 |
| WP | AAI99683_01 | 100001 | 210000 |
| WP | AAI99683_02 | 200001 | 310000 |
| WP | AAI99683_03 | 300001 | 410000 |
| WP | AAI99683_04 | 400001 | 510000 |
| WP | AAI99683_05 | 500001 | 610000 |
| WP | AAI99683_06 | 600001 | 710000 |
| WP | AAI99683_07 | 700001 | 810000 |
| WP | AAI99683_08 | 800001 | 910000 |
| WP | AAI99683_09 | 900001 | 1010000 |
| WP | AAI99683_10 | 1000001 | 1110000 |
| WP | AAI99683_11 | 1100001 | 1210000 |
| WP | AAI99683_12 | 1200001 | 1310000 |
| WP | AAI99683_13 | 1300001 | 1410000 |
| WP | AAI99683_14 | 1400001 | 1510000 |
| WP | AAI99683_15 | 1500001 | 1610000 |
| WP | AAI99683_16 | 1600001 | 1710000 |
| WP | AAI99683_17 | 1700001 | 1810000 |
| WP | AAI99683_18 | 1800001 | 1910000 |
| WP | AAI99683_19 | 1900001 | 2010000 |
| WP | AAI99683_20 | 2000001 | 2110000 |
| WP | AAI99683_21 | 2100001 | 2210000 |
| WP | AAI99683_22 | 2200001 | 2310000 |
| WP | AAI99683_23 | 2300001 | 2410000 |
| WP | AAI99683_24 | 2400001 | 2510000 |
| WP | AAI99683_25 | 2500001 | 2610000 |
| WP | AAI99683_26 | 2600001 | 2710000 |
| WP | AAI99683_27 | 2700001 | 2810000 |
| WP | AAI99683_28 | 2800001 | 2910000 |
| WP | AAI99683_29 | 2900001 | 3010000 |
| WP | AAI99683_30 | 3000001 | 3110000 |
| WP | AAI99683_31 | 3100001 | 3210000 |
| WP | AAI99683_32 | 3200001 | 3310000 |
| WP | AAI99683_33 | 3300001 | 3410000 |
| WP | AAI99683_34 | 3400001 | 3510000 |
| WP | AAI99683_35 | 3500001 | 3610000 |
| WP | AAI99683_36 | 3600001 | 3710000 |
| WP | AAI99683_37 | 3700001 | 3810000 |
| WP | AAI99683_38 | 3800001 | 3910000 |
| WP | AAI99683_39 | 3900001 | 4010000 |
| WP | AAI99683_40 | 4000001 | 4110000 |
| WP | AAI99683_41 | 4100001 | 4210000 |
| WP | AAI99683_42 | 4200001 | 4310000 |
| WP | AAI99683_43 | 4300001 | 4403765 |
| Alignment Scores: | | | |
| Pred. No.: | 9.69e-12 | Length: | 110000 |
| Score: | 685.00 | Matches: | 530 |
| Percent Similarity: | 32.37% | Conservative: | 186 |
| Best Local Similarity: | 23.96% | Mismatches: | 852 |

| | | |
|------------------------|----------|-------------------|
| Alignment Scores: | | |
| Pred. No.: | 9.69e-12 | Length: 110000 |
| Score: | 685.00 | Matches: 530 |
| Percent Similarity: | 32.37% | Conservative: 186 |
| Best Local Similarity: | 23.96% | Mismatches: 852 |

QY 1582 AspArgLysLeuThrSerThrProArgGluile-----AlaLysSerPro 1596
Db 17192 CGACGGCTACCATCACATCAACCCCTACTACTACGCGGACACTCCCTCGCGGAATCACCG 17251
QY 1597 -----HisSerThrValPro-GluHisHisProHisProHisSerPr 1610
Db 17252 CGCCCACTCGCGGATCTCTCACTCTACCGACGCGCAACCCCTCATCACCGCGG 17311
QY 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisileProLe 1630
Db 17312 CMACCTCATGCAACATGCGCGCGCGGACCATGACCCCTCCACACACCCCGCAC 17371
QY 1630 uAlaPheAspProThrSerileProArgGlyileProLeuAspAlaAlaAlaTyr-- 1649
Db 17372 ACATCACCCACCACTCACCGCGCGCAACGACCTCGCGCATCACACACCC 17431
QY 1650 -TyrLeuProArgHisLeuAla----- 1656
Db 17432 CCACCTCCCTCGTCATCAGCGGCGACCCCGCCACCGCTCCACACATCACACCCCTCTGCC 17491
QY 1656 ----- 1656
Db 17492 AACACAGGCATCAAAACCAAAACCCCTCCCGACCAAAAGCGCTTCCACTCCCGCCACA 17551
QY 1657 -----ProAsnProThr-----TyrProHisLeu-TyrProProT 1668
Db 17552 CCAACCCCTCTCAACCACTCCACGAGCACACCAACCCCTCACTACACCCCGACCC 17611
QY 1668 yr-----LeuileArgGlyTyProAspThrAlaAlaLeu----- 1679
Db 17612 ACACCCCTCATCACCGCGCAACACCCCGCCGACCACTCTCTCACCCCGCCACTCTGGA 17671
QY 1680 -----GluAsnArgGlnThrIleleAsnAspTyrlleThrSerGlnGlnMetHis- 1697
Db 17672 CCAACAGCGCGCACACCGTC-----GACTACGCGCACCAACCCCGCAACCCCTCCACC 17725
QY 1698 AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu 1717
Db 17726 AACACGGCGTCACCACTACATCGAAGTCGAC-----CCGACACAA 17767
QY 1718 SerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleleAspLeuSerGlnVal 1737
Db 17768 CCCTCA----- 17773
QY 1738 ProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArg 1757
Db 17774 CCACCTCACCCACCACTCCCAACACCCCGCCGACCACTCCCTCACCTCACCCACC 17833
QY 1758 LeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeu 1777
Db 17834 CCCACC-----ACCACCCCAACACCCCTCTCTCAACACCTCGCCCAAAACCA 17884
QY 1778 SerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlu 1797
Db 17885 CCACCTTGCGACCCCGAC-----ACTACACCCCGGACCAACCAACCAACCC-CAACCCAC 17937
QY 1798 ArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSer 1817
Db 17938 ACCACCTCGAC-----CTCCCC 17955
QY 1818 ThrThrThrValGluHisAlaProIleTyr-----ArgProGlyThrGluGln 1833
Db 17956 ACCTACCCCTTCAACACACGACTACTGCTCGAAAGCACACAGCGGCTGCGGATCC 18015
QY 1834 SerSerGlySer--SerGlySerSerGlyGlyGlySerSer----- 1848
Db 18016 GGTTCGGGTTCCGGTTCGGGCGGAGGACTCGCGGCGGAGCGGACGAGAGGTGGAGTCG 18075
QY 1848 ----- 1848
Db 18076 CGGTTTGGGACGGGTGGCGCGGACCTGGAAACGGTCGCGGACCACTCGCGGTG 18135

QY 1849 -----SerArgProAlaSerHisSerHisAla 1857
Db 18136 CCCCCCTCGCGCGCTGGACACGGTGGTGGCGCACTCTCCGCTGGCACCGCCACCA 18195
QY 1858 HisGln----- 1859
Db 18196 CAGCAACAGCGCGCATCAACACTCGGACCTACCGAAACCTGGAAACCCCTCACCTC 18255
QY 1860 -----His-SerProileSerProArgThrGlnAspAlaLeuGlnGlnArgProse 1876
Db 18256 CCAACACCCACCAACCCCGCAACCTCGCTCATCGCCATCCCGGAAACCCAGACCCAC 18315
QY 1876 rValLeuHisAsnThrGlyMetLysGlyIleleThrAlaValGluPro---SerLysPr 1895
Db 18316 C-----ACCCCGCATCATCAACCATCTCTCCAC 18342
QY 1895 oThrValLeuArgSerThrSerSerProValArgProAlaAlaThrPheProPr 1915
Db 18343 AAC-----TCACACACACGCGCATCACCCCTCCCTCCCTCAACACACAC 18393
QY 1915 oAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluPr 1935
Db 18394 C---ACACCAACCCCAACACCTCCACACCGACCAACAGCCCAACACACACAC 18450
QY 1935 oValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspTh 1955
Db 18451 GGACCCATCACCGGCTGCTCTCTCTCGCGCTCGAGGAAACACACCCCGCCACCC 18510
QY 1955 xGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPr 1975
Db 18511 CCCC-----ACACACCCCGCGACCTCTCTCAACCTCACCTCACCCCAAC 18558
QY 1975 oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl 1995
Db 18559 CACA-----CCCAACCCCGCCACCAACCCCTCTCTGAGCGCCACCAAC 18606
QY 1995 aArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro----- 2009
Db 18607 GCACACCAACCCCGCAACCGACCCCTCAACACCCCGCAACCGCCCAACCTGGGA 18666
QY 2010 -----AspProProAlaProProAlaSerAlaSerAspPr 2021
Db 18667 CTGCGCGGACCACTCTCTCGAACACCCCGCCCGCAACCGCGGATCATCGACCTC-CC 18725
QY 2021 O-----HisArgGluLysThrGlnSerLysProPheSe 2032
Db 18726 CACACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 18785
QY 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
Db 18786 CCAACCCCAACTCGCATCGCACCGCACCGCACCGCACCGCGCGCGCTCATCGCCCGCC 18845
QY 2052 yValGluProValSerProValSerProSer-----LeuThrHi 2066
Db 18846 CTTACCCCGCCACACACACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 18905
QY 2066 saePlyGlyLeuProLysHisLeuGluLeuAspLysSerHis----- 2081
Db 18906 CGGAACCGGCGC-CTTGGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 18964
QY 2082 -----LeuGluGlyGluLeuArgProLysG 2090
Db 18965 AACACCTCTCTCTCACGACCGGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 19024
QY 2090 nProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuPr 2110
Db 19025 CCAACTCCAAACAAAA-----AGGCATCCACCTCACCATCACCACTCGCG 19069
QY 2110 oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisG 2130
Db 19070 ACACGACCAACCGACCACTCCA-----ACACTCTCT-----CAACACCA 19111
QY 2130 nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi 2150

[illegible]

| | | | |
|----|-------|---|-------|
| Qy | 563 | laValalaSerIysGlyArGLysThrAlaAsnSerGln-----GlyArgArgL | 579 |
| Db | 13046 | TCGTCGAAGCCTACGGCTCAGGACCGCCCAACGGCGCCCTCTGGCTCGGAACCCCTCA | 13105 |
| Qy | 579 | ysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer----- | 592 |
| Db | 13106 | AGTCCAACATCGCGCACTCCATGGCGCGCTGCGGGGTGCATCAAGATGGTGA | 13165 |
| Qy | 593 | -----GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuA | 610 |
| Db | 13166 | TGCGCGCTGCGGAATGCTGCTCCCGCG-----ACGTTGCATGTGG | 13207 |
| Qy | 610 | enGluSerSer-----ArgTTPThrGluGluGluMetGluThrAlaIysIysGlyL | 627 |
| Db | 13208 | ATGAGCGCTCGCGCATGTGACTGGTCCGCGGTGCGGTGCAGCTG-----C | 13255 |
| Qy | 627 | euleuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerIysThrVal- | 646 |
| Db | 13256 | TGACGAGAC-GGT-----GCCCTGGCGCGGAGGGCGGTACGCGCG | 13302 |
| Qy | 647 | -----SerGlnCysLysAsnPheTyr=PheAsnTyrLysLysArgGlnAsnLeuAspGluI | 665 |
| Db | 13303 | GCAGGAGTGCATCTTCGGGTACCGGCACCAAGCCCACTCATCTCTCGAGGAAGCA | 13362 |
| Qy | 665 | leLeuGlnGlnHisLysLeuMetGluLysGluArgAsnAlaArgArgIysLysLysL | 685 |
| Db | 13363 | CCGCGCCACAACATCCGTCAGACACACCCGCGCAGCGTCCCGGAGATCAGCCGCC | 13422 |
| Qy | 685 | ys-AlaProAlaAlaSerGluGluAlaA-----PheProPro | 698 |
| Db | 13423 | GACGAGGATCCGGTAGTGGCATGAGGTGTCTGCCGCGAGTCCAGGGGTGTGGCGTGG | 13482 |
| Qy | 699 | valValGluaspGluMetGluAlaSerGlyValSerGlyAsnGluGluMetVal | 718 |
| Db | 13483 | CTGCTGTGCGCCAAAGTGCAGCGCGCC-----CTGCGC | 13515 |
| Qy | 719 | GluGluAlaGluAlaLeuHisAla-SerGlyAsnGluValPro-----ArgGlyGluCy | 736 |
| Db | 13516 | GCCAGGCGCCAGCCCTCGACGCCCACTCACCGCACCCCGCGCTCGACCTCGCCGAC | 13575 |
| Qy | 736 | sSerGly-----ProAlaThrValAsnAsnSerSerAspThrGluSerI | 751 |
| Db | 13576 | GTGCGGTACACCTCGCCACGCGCGCGCGTGTTCGACCAACGCGGCCCACTCATCGCC | 13635 |
| Qy | 751 | ePro-----SerProHisThrCluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr | 769 |
| Db | 13636 | GCCGACGCGACACCTCTCTGAAGCACTCTCAAGGCACTCGCGCAGGCGAACCACCC | 13695 |
| Qy | 769 | oProAlaThrLeuGlyAlaAsp-----GlyProProGlyProPro----- | 783 |
| Db | 13696 | GCGCTATCCACAGACGCCCCAGCGGGGACCGGACCGGGAGCGCGAGAAAGACC | 13755 |
| Qy | 784 | -----ThrProProArgThrSerArgAlaProIleGluProThr----- | 797 |
| Db | 13756 | GCATTATCTGCTCCGACAGGGCACCAACGCGCGCATGGCCACGCGCTCTACCAC | 13815 |
| Qy | 798 | -----ProAlaSerGluAlaThrGlyAlaPro----- | 806 |
| Db | 13816 | ACCCACCCCGTCTTCGCCCGCGCACTCAACGACATCTGCACCACTCGACCCCACTC | 13875 |
| Qy | 807 | -ThrProProAlaProProSerProSerAlaProProValValProLysGluG | 826 |
| Db | 13876 | GACCACCCCTCTCTCCCTCTCTACCCAGGACCCCAACACCAAGACACACACCCCTC | 13935 |
| Qy | 826 | uLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProPr | 846 |
| Db | 13936 | GAAGAAGCGCGCACTGCTCCAGACAGACCCGCTACGCGCCAGCCGCTCTTCGCGCTT | 13995 |
| Qy | 846 | oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerG | 865 |
| Db | 13996 | CAGTCCGCTCCACGCGCTCTCTCACCGACGGCTACCATCACCCCACTACTACGCC | 14055 |

| | | | |
|----|-------|--|-------|
| Qy | 865 | uCyThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAaGluAl | 885 |
| Db | 14056 | GGACATCCTCGCGGAAAATACCGCGGCCACCTCGCGGCATCTCACCTCACCAC | 14115 |
| Qy | 885 | aThrAlaGluGlyAlaLeuLysAlaGluLysLys | 896 |
| Db | 14116 | GCCACCACCTCATCACCCGCGCCACCTCATGCAACCATGCCCGCGGCACCATG | 14175 |
| Qy | 897 | -GluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSer | 916 |
| Db | 14176 | ACCACCCTCCACACACCCCCACACCATCACCCACCTCACCGCCCGCAAAACGAC | 14235 |
| Qy | 916 | pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsn | 936 |
| Db | 14236 | CTCGCATCGCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCA | 14284 |
| Qy | 936 | gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAla | 956 |
| Db | 14285 | -----CCCCCAGCCGTCGAACATCACCA | 14328 |
| Qy | 956 | rProGlnLysProLeuAspLysGlnLeuLysGlnArgAlaAlaLysPro | 975 |
| Db | 14329 | GGCATCAAAACCAAAACCTCCCAACCAACGCGCTTCCACTCCCCCCACCAACCC | 14388 |
| Qy | 975 | oIleGlnValThrLysValHisGluPro | 994 |
| Db | 14389 | ATCCTCAACCAACTCCACGACACACCCAAACCTCACCTACCAACCCACCAACCC | 14448 |
| Qy | 994 | oAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProThrLys | 1014 |
| Db | 14449 | CTCATCACGCGCAACACCCCGCAACCACTCTCACCCCCCACTACTGGACCCCAAA | 14508 |
| Qy | 1014 | nProGlySerSerProArgLysSerArgSerProAlaProProAlaAspLysGluAl | 1034 |
| Db | 14509 | GCCCGCAACA | 14556 |
| Qy | 1034 | aPheAlaAlaGluAlaGlnLys | 1048 |
| Db | 14557 | CACGGCGTCACCACTACATCGAACTCGGACCCGACACACCTCACCACTCACCCAC | 14616 |
| Qy | 1048 | sTrpThrSerGlyLeuProPheProValPro | 1063 |
| Db | 14617 | GACAACCTCCCCAACACCCCGCAACCACTCACCTCACCCCGCCACCAACCC | 14676 |
| Qy | 1063 | eLysAlaSer | 1082 |
| Db | 14677 | CAAAACCCCTCTCTCAACCTCGCAAAACCAACCACTCGGACCCCAACCACTAC | 14736 |
| Qy | 1082 | iSProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProThrI | 1102 |
| Db | 14737 | ACCCACCA | 14787 |
| Qy | 1102 | leSerAsnPro | 1117 |
| Db | 14788 | TACCCCTTCCAAACACCACTACTGGCTTCCAAACCAACCGCGCAACCGAGCAC | 14844 |
| Qy | 1117 | alLeuGluArgGln | 1122 |
| Db | 14845 | TCACCGAGGAGCGCGTGAGCGACCAACCCCATCAACCCCGCTGGTGATGCTCTC | 14904 |
| Qy | 1122 | leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTySerGluHisAla | 1141 |
| Db | 14905 | GTGGGCAAGTCTCCGAGGAGCGGACGAGAGCTGTCGCGCTGGTGGCACCATGG | 14964 |
| Qy | 1142 | LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu | 1161 |
| Db | 14965 | GCCGCTGTGTGGGCCATGCCACTCCCGAAGTGATGCTTCCGAACAAGGCTTCAAGAG | 15024 |
| Qy | 1162 | AlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro | 1180 |
| Db | 15025 | CTGGGTTTTGATCTCTCCCGCAATTCAGCTTCGTAATCGACTCTGCTGCGCTGAC | 15084 |
| Qy | 1181 | -----ProGluSerLeu | 1194 |

Alignment Scores:
 Pred. No.: 3,63e-12 Length: 30690
 Score: 685.00 Matches: 704
 Percent Similarity: 30.18% Conservative: 300
 Best Local Similarity: 21.16% Mismatches: 1193
 Query Match: 5.18% Indels: 1144
 DB: 4 Gaps: 148

US-09-522-753-5 (1-2517) x AAH79277 (1-30690)

Qy 44 HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleLeuGlnProGlnArg 63
 Db 11193 CATCATCGTTCGGTCTGATCCATCCATGTCACCCCGA-----CGC 11234
 Qy 64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluHis 83
 Db 11235 ACTGGCCCTCTTCGATGGGC-----CTGGCTCGACCGCGCGTCTGCTGCC 11285
 Qy 84 LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPhe11 103
 Db 11286 CGCGACGCTGCTCCCGCCCGCCCTGCGCCCTG-----CTGCAGGACCTCCT 11336
 Qy 103 eGluSerLysArgProGluLeuLeuLeu-ProAspProLeuLeuArg---ProSerP 122
 Db 11337 GCCCGCACCCCGCGCGCACACCGCACACCATACCGGTGGTGGGACACGCGC 11396
 Qy 122 roLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuT 142
 Db 11397 CCAGTCGACCCCGCTGGCGCGCCGACACACGACACACACACCCCTCTCTCGC 11456
 Qy 142 hrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeuGlu 162
 Db 11457 CCGTGTCC---GCTCCCATCGCACCGCTCTG-GGCCACACACCCCCC-----GACA 11506
 Qy 162 euValProProArgLeuSerLysGluGluLeuLeuGlnAsnMetAspArgValAspArg 182
 Db 11507 CCATCCCCCCC-----GACCGCGGTTCGCG 11533
 Qy 182 luileThrMetValGluGlnLeuSerLysLeuLysLysGlnGlnGlnLeu-Glu 201
 Db 11534 ACCTCGGCTCGACTCCCTCCACCGCGCTCGAATACGACACCGCTCTCCCGCACACCG 11593
 Qy 202 GluGluAlaAlaLysProProGluProGluLysProValSerProPro-----ProIle 219
 Db 11594 GACTCGGCTCCCGACACCCCTCGCTCGACACCCACCCACCCACCCCTCACCCACC 11653
 Qy 220 GluSerLysHisArgSerLeuValGlnIleLysAspGluAsnArg-LysLysAlaGl 239
 Db 11654 ACCTCCACACAACTCC-----AGCCACACCGGACAAACGCTGT-- 11693
 Qy 239 uAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu---ProLeuTyrAs 258
 Db 11694 -----CGCCCGCTGTGGCGGAGCTCGACAAACTCGAATTCGCCCTCTC----- 11738
 Qy 258 nGlnProSerAspThrArgGln-----TyrHi 267
 Db 11739 -CGCCCTCGACAAACCGACACCGCGCGGAGAGTCACTCGGCTGAAGTCACTCA 11797
 Qy 267 sGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuLeuLeuTyrPheLysArg 287
 Db 11798 TGTGAGGTGGAACG-CACCCAGCATCGACAGCGGAAAGCGCTGATGACGACGAGAAG 11856
 Qy 287 gAsnHisAlaArgLysGlnTrpLysGlnLysPhe----- 298
 Db 11857 TTCACAT-----CGCAACAGAGCTGAGATTTCAATTCAATTGACACACCTCGC 11910
 Qy 299 -CysGlnArgTyrAspGlnLeuMetGluAlaLeuLysLysValGluArgIleGluAs 318
 Db 11911 CTGTCTGACCGGACCGCTGCCACTCCGCC-----CGTATCCGCTG 11952
 Qy 318 nAsnProArgA-gArg-----AlaLysGluSerLysValArgGlyTyr-Gl 334
 Db 11952

Db 11953 -GGCCCTCTAGGACGTGAATGCAATTGGCGAATGAAGCAAGCTCTCTGGAATACCTCAA 12011
 Qy 334 uLysGlnPheProGluIleArgLysGlnArg-----GluLeuGlnGluArgMe 350
 Db 12012 GCGGCTACTCGGACCTGGACCGACTCGCGTGCCTGTACGAGGTGGTTCGAGCGTGA 12071
 Qy 350 tGlnSerArgValGlyGlnArgGly-----SerGlyLeuSerMetSerAl 365
 Db 12072 GCAGGACCGATCGCATTTGTGGGATGCGGTGCTTACCCAGCGCGGCGACGTCACC 12131
 Qy 365 aAlaArgSerGluHisGluValSerGluIleLeuAspGlyLeuSerGluGlnGlu----- 383
 Db 12132 CACGCGACTGTGCGCATCTCGTCAAGTCCAGACGCGATTCGGGAGATTCGCCACCGA 12191
 Qy 384 -----AsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
 Db 12192 CCGTGATGGAACCTGGACGAG-----CTCTA 12218
 Qy 400 rAspAlaAspGlnGlnArg-----IleLysPheLeuAsnMetAsnGlyLeuMetAlaAs 418
 Db 12219 CGACCGGACCGCGCTCAGGAACAGTTACACGCGCAGCGGAGGTTCTCTATGA 12278
 Qy 418 p-----ProMetLysValTyrLysAspArgGlnValMetAsnMetTr 432
 Db 12279 CGCGGCGACTTCGACGCGCGTTCCTCGAGTTGTCCCGGTGAGCGCTGGCAATGA 12338
 Qy 432 pSerGluGlnGlu-----LysGluThrPheArgGluLysPheMetGln----- 446
 Db 12339 CCGCGACGACGCGCTCTGCTCGAAACACTTGGGAAACGTTTCGAACAGCGGGAATCGA 12398
 Qy 447 -----HisProLysAsnPh 451
 Db 12399 CCGGAGGTCCATCGCGGGAAGCGGACCGGGTTCCTGGGGATCAATCCGAGGACTA 12458
 Qy 451 e-----GlyLeuIleAlaSerPheLeuGluArg-----Lys 461
 Db 12459 CACACCGGATACAC-ACATCAGCCCTCAACACGAGTCGAGGCTACCTGCTCAGTGGCA 12517
 Qy 461 sThrValAlaGluCysValLeuTyrTyrTyrLeu-----ThrLysLysAsnGl 477
 Db 12518 GCGCGGACGATTCGCTGAGCGGTATCTCTTCACTTCACTTGGGCTCGAAGGCGCTCGCA 12577
 Qy 477 uAsnTyrLysSerLeuVal----- 483
 Db 12578 TCATATCGACACCGCGTGTCTCTCGCTCGCTCGCCCTGCAATCTGCGCTGCAAGCGC 12637
 Qy 484 -----ArgArgSerTyrArgArgArgGlyLysSerGl 494
 Db 12638 TCCGCTCCGTGAATGACCATGCGCTCGACGCGCGCTCGCTCATGCGCACCTCCCT 12697
 Qy 494 nGln----- 495
 Db 12698 TCGTCTTACCGAGTTCTCTCGCAGCGGCGCTGCGCGCAGCAGCGCGCTGCAAGCGCT 12757
 Qy 495 ----- 495
 Db 12758 TTTGCGGCGCGCGGACGCGGACCGGCTGTCGAGGGTGTGGGATGCTCTGCTGTGAGC 12817
 Qy 496 -----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
 Db 12818 GGCTCTCGAGCGCGCGGCAACGCTCAGCTGCTCTGCGCGCTGCTCGCGGACGCGCGC 12877
 Qy 509 nGlnProMetProArgSerSerGlnGluLysAspGluLysGluLysGluLysGluLys 529
 Db 12878 TCACACGAGGCGCGGCAACGCGCTGACCGCACCCACCGCTGCTTCAAGTCAAGG 12937
 Qy 529 aGluLysGluGluLysPro-GluValGluAsn-----AspLysG 543
 Db 12938 T-----CATCGCGCGCTTGGCCACGACACCTCTCCCTCTCCGCTGCGCATGTCG 12985
 Qy 543 luAspLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGlu 563
 Db 12986 ATGCGGTGGAGGCCACCGGACGCGGACCCCTTGGGCGACCCCGATCGAGCTCAAGCC 13045

Db 40740 GCCGCGTTCGCCGCGTTCAGCATCAGACGGCGCTGCCCGCCAGCGCGCCCTTCGCGCC 40681
Qy 760 -----LysAspThrGlyGlnAsnGly-----ProLysProAlaThrLeuGlyAl 775
Db 40680 GATGTGGAAGACCGGACAAACCGTGGCGCGCGCGCGCGCGCTACCTCCGAGACC 40621
Qy 775 aAspGlyPro-----ProProGlyProProThrProPro-----ArgArgThrSerAr 791
Db 40620 GAACGCTCCGAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40561
Qy 791 gAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProAl 811
Db 40560 GCG 40504
Qy 811 aProProSerProSerAlaPro-----ProProValValProLysGluGluLy 827
Db 40503 CCG 40444
Qy 827 sGluGluGluThrAlaAlaAlaProProValGluGluGluGluGluGluGluGlu 847
Db 40443 GCCGAAGAGCAAGCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40408
Qy 847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysTh 867
Db 40407 GACCACCG 40375
Qy 867 rGluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluAl 887
Db 40374 -----GCTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40322
Qy 887 aGluGluAlaLeuLysAlaGluLysLysGluGlySerGlyArgAlaThrThrAlaLy 907
Db 40321 CGGCG 40262
Qy 907 sSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAlaAspGluValAs 927
Db 40261 TGGCG 40202
Qy 927 pGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuThrPr 947
Db 40201 TGCTTCCCG 40142
Qy 947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeu 967
Db 40141 CCGCGCGCGATCCCGCAGCCAGCGGGGCTGCGCAACACCACCGG-CGCGCGCGCTC-- 40085
Qy 967 sGlnArgAlaAlaAlaThrProProlleGlnValThrLysValHisGluProArgGl 987
Db 40084 -CCCCGTG-GCGAACAGCGCGCGGTTA-----CGCGCGCGCGCG 40048
Qy 987 uAspAlaAlaProThrLysProAlaProAlaProProProGlnAsnLeuGlnPr 1007
Db 40047 GCCATTGCCACCG 40008
Qy 1007 oGluSerAspAlaProGlnGlnProGlySerProArgGlyLysSerArgSerProAl 1027
Db 40007 -----GCCCG 39967
Qy 1027 aProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPr 1047
Db 39966 CCGGCCATTGCCACCG 39907
Qy 1047 oCysTrpThrSerGlyLeuProPheProValProProArgGluValLysAlaSerPr 1067
Db 39906 GAGCAGCGCGCTCCG 39847
Qy 1067 oHisAlaProAspProSerAlaPheSerTyAlaProProGlyHisProLeuProLeuGl 1087
Db 39846 GCTGCCCGCGGTGCGCGCGGTGCGCAACAAGATCCCGCGCGCGCGCGCGCGCGCGCG 39787
Qy 1087 y-----LeuHisAspThrAlaArgProValLeuLeuProArgProThrIleSerAnPr 1105
Db 39786 CCGCGCGGTGACGCGCGCGATCGTTCTTCCGCGCGCGGTTCCTCCGCGCGGTTCGCGAACCA 39727

Qy 1105 oProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaI 1125
Db 39726 GCGCGCGCGTCCGCGCTCAGCCCGCGTCCCGAGTGCC-----GTTGGCGCGCT 39677
Qy 1125 eSerGln-----GlyMetSerValGlnLeuHisValProTrSerGluHisAl 1141
Db 39676 TGCCGATCAGCGACGCGCGTGTGGAAAGGGGTGTTTACCATTATTCACGCT 39617
Qy 1141 a----- 1141
Db 39616 CTTGCTGAGGGTGTCTCAGGGGCGAGTGTCTGCGGGGGCGCTAAAGCGCTCCAGGCCCA 39557
Qy 1142 -LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLe 1161
Db 39556 ACAGACCCCGCGGTACCGCGATCCGCGCTTCCAGGGTCCGCGCGTCCGCGCG 39497
Qy 1161 uAlaProPheSerGlyValLysGlnGluLeuSerProArgGlyGlnLaglyProPr 1181
Db 39496 TGCGCGCTGTCGCGCGTTCGCGATCAACACGCGGTGCGCGCGCGCGCGCGCGCAC 39437
Qy 1181 oGluSerLeuGlyValProThrAla--GlnGluAlaSerValLeuArgGlyThrAlaLe 1200
Db 39436 C-----GTCCCGCGCGCTGCTCCGCGCGCGCTTCCGCGCGTCCGCGCGCT 39389
Qy 1200 uGlySerValProGlyGlySerIleThrLys-----GlyIleProSerThrArgVa 1217
Db 39388 TGCCGATCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCCCGACGCTTTTAC 39329
Qy 1217 lProSerAspSerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspVa 1237
Db 39328 TAATGCG 39274
Qy 1237 lLeuTyLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspAr 1257
Db 39273 -----GCCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39242
Qy 1257 gGlyArgGluAspSerLeuProLysGlyHisValIleTyArgGlyLys----- 1273
Db 39241 TGCGACTAGCGCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39182
Qy 1274 ----LysGlyHisValLeuSerTyArgGlyGlyMetSerValThrGlnCysSerLysGl 1292
Db 39181 TGCG 39123
Qy 1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyAs 1312
Db 39122 CTTCCGCGCGCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39065
Qy 1312 pMetMetGluGlyArgValGlyAlaIleSerSerAlaSerIleGluGlyLeuMetGl 1332
Db 39064 ----CGCGCTTCGAGAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39009
Qy 1332 yArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisLeAr 1352
Db 39008 GCGTTCG 38955
Qy 1352 gGlySerIleThrGlnGlyIleProArgSerTyValGluAlaGlnGluAspTyLeuAr 1372
Db 38954 G-----CG 38916
Qy 1372 gArgGluAlaLysLeuLysArgGluGlyThrProProProProProProProProPro 1392
Db 38915 CCGCGCTTACCG 38865
Qy 1392 pLeuThrGluAlaTyLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGl 1412
Db 38864 -----CGTTG----- 38859
Qy 1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGl 1432
Db 38858 -----CGATCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38833

[illegible]

RESULT 28

AAI99682_08/c
Continuation (9 of 45) of AAI99682 from base 800001 (Mycobacterium tuberculosis strain H37Rv)
Up sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | AAI99682_00 | 1 | 110000 |
| WP | AAI99682_01 | 100001 | 210000 |
| WP | AAI99682_02 | 200001 | 310000 |
| WP | AAI99682_03 | 300001 | 410000 |
| WP | AAI99682_04 | 400001 | 510000 |
| WP | AAI99682_05 | 500001 | 610000 |
| WP | AAI99682_06 | 600001 | 710000 |
| WP | AAI99682_07 | 700001 | 810000 |
| WP | AAI99682_08 | 800001 | 910000 |
| WP | AAI99682_09 | 900001 | 1010000 |
| WP | AAI99682_10 | 1000001 | 1110000 |
| WP | AAI99682_11 | 1100001 | 1210000 |
| WP | AAI99682_12 | 1200001 | 1310000 |
| WP | AAI99682_13 | 1300001 | 1410000 |
| WP | AAI99682_14 | 1400001 | 1510000 |
| WP | AAI99682_15 | 1500001 | 1610000 |
| WP | AAI99682_16 | 1600001 | 1710000 |
| WP | AAI99682_17 | 1700001 | 1810000 |
| WP | AAI99682_18 | 1800001 | 1910000 |
| WP | AAI99682_19 | 1900001 | 2010000 |
| WP | AAI99682_20 | 2000001 | 2110000 |
| WP | AAI99682_21 | 2100001 | 2210000 |
| WP | AAI99682_22 | 2200001 | 2310000 |
| WP | AAI99682_23 | 2300001 | 2410000 |
| WP | AAI99682_24 | 2400001 | 2510000 |
| WP | AAI99682_25 | 2500001 | 2610000 |
| WP | AAI99682_26 | 2600001 | 2710000 |
| WP | AAI99682_27 | 2700001 | 2810000 |
| WP | AAI99682_28 | 2800001 | 2910000 |
| WP | AAI99682_29 | 2900001 | 3010000 |
| WP | AAI99682_30 | 3000001 | 3110000 |
| WP | AAI99682_31 | 3100001 | 3210000 |
| WP | AAI99682_32 | 3200001 | 3310000 |
| WP | AAI99682_33 | 3300001 | 3410000 |
| WP | AAI99682_34 | 3400001 | 3510000 |
| WP | AAI99682_35 | 3500001 | 3610000 |
| WP | AAI99682_36 | 3600001 | 3710000 |
| WP | AAI99682_37 | 3700001 | 3810000 |
| WP | AAI99682_38 | 3800001 | 3910000 |
| WP | AAI99682_39 | 3900001 | 4010000 |
| WP | AAI99682_40 | 4000001 | 4110000 |

WP AA199682_41 4100001 4210000
 WP AA199682_42 4200001 4310000
 WP AA199682_43 4300001 4410000
 WP AA199682_44 4400001 4411529

Alignment Scores:
 Pred. No.: 7,31e-12 Length: 110000
 Score: 689.00 Matches: 526
 Percent Similarity: 33.44% Conservativeness: 207
 Best Local Similarity: 24.00% Mismatches: 872
 Query Match: 5.21% Indels: 599
 DB: Gaps: 91

US-09-522-753-5 (1-2517) x AA199682_08 (1-110000)

QY 546 LeuLysGluLysThrAspThrSerGlyGluAsnAspGluLysGluAlaValala 565
 DB 41474 GTTCGTCGGAAGTGTATGGCGTGGCGGTGACGGCTTCGACGAATCGCATCTGCG 41415
 QY 566 SerLysGlyArgLysThrAlaLanSer-----GlnGlyArg-- 578
 DB 41414 TCGCGGTGACGGAATCTCGAAGATCGGTGCGATCGTTGCCAGCCGGAGACGACGCCCA 41355
 QY 579 -----LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
 DB 41354 TACCAGTTCCGCACCTGTGAGGGGTCTGTCAGCCGGAGCAGTCGATCGAACACGAGGCGG 41295
 QY 593 Glu-----GluAlaIleThrProGlnGln----- 600
 DB 41294 GACGGTTTCGGTGATTTCGGGTGGTCACCGCGGTGTGCAGCACGACACCGTTGGCGTTCGAG 41235
 QY 601 -----SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrrp 615
 DB 41234 AAGGAACATTTCGCGTCTTCGCGCTATTTGGGGTTACTTCGGGAGTTTCAGTTCACAGGCC 41175
 QY 616 ThrGluGluGluMetGluThrAla----- 623
 DB 41174 TCGTCGACACTTCGGACACGCTCTATTTCGAGGTTCAGTTCGATCGCTCGCCGCGGACCG 41115
 QY 624 -----LysLysGlyLeuLeuGluHisGlyArg-----AsnTrpSerAla 636
 DB 41114 GTGCGCGCGTTCGAAACCGGACGGTC-----GGCGGGCGCCGCGACGTGGCGGCGG 41061
 QY 637 Ile----- 637
 DB 41060 GCGAACTCCCGGACGAGGGCTCTCGATCAGCCGCCAGCGATTGACCATCGCTCGCGG 41001
 QY 638 AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys 657
 DB 41000 GCCCGCGCTTTGGCGGACGCGAGTATTCTCA----- 40971
 QY 658 LysArgGlnAsnLeuAspGluIleLeu-----GlnGlnHis 669
 DB 40970 -----TCGGAGATTGACACCGTGTGCGCATGATGCACCGGTAAACGATCAACAGCAT 40917
 QY 670 LysLeuLysMetGluLysGluArgAsnAlaArgLysLysLysLysAlaProAlaAla 689
 DB 40916 CATCTCGCGTCT-GAGCGGCCGACGCGATGCTGTCAACGCCGCGCGCGGTGCGAGGCCCGC 40858
 QY 690 AlaSerGluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGly 709
 DB 40857 CGGCTACATCAAGCGGTTTCAGCCCGGTTCTCACCGAGCAGCAGCCGCGCGCGCGCGG 40798
 QY 710 ValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsn 729
 DB 40797 GCCGTGGG----- 40788
 QY 730 GluValProArgGlyGluCysSer-----GlyProAlaThrValAsnAsnSer 745
 DB 40788 -----GCCGGGTGCACCCCGGATTTCGCCGCGTTACCGGTGTTACCGCGGTT 40741
 QY 746 SerAspThrGluSerIleProSerProHisThrGluAlaAla----- 759

```
QY 2195 gGlySerProHis-----SerGluGlyG1 2203
Db |||||
9561 GTCCATCGCACACCCGCAACGCGGACACCGCCATCACCACCCACCCCTGC 9502
QY 2203 yLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyLeG1 2223
Db :::::::::::
9501 CGAGAAACTCCACAAACACCCCGCG----- 9474
QY 2223 uProValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrPr 2243
Db |||||
9473 -CCGCAATCACTGTCACCCCGGCAACGCGCATCATCACCACCCACCCGCAACCC 9415
QY 2243 oLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProG1 2263
Db |||||
9414 T-----GACAGCGCAATGCAAGCCACCAACGACGACGACGCGGTATCAACC 9364
QY 2263 yAsnThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
Db ::::::|
9363 GAAATCGACGACCTCTCAAC----- 9342
QY 2283 lLysSerLysGlnGluLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluPro-- 2302
Db |||||
9341 -----CACAACATACGAAACGACCGCGAC 9316
QY 2303 -GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
Db ::::::|
9315 ACCACACTCGCGGACGTCGCGTCAACA-----ATAACCT----- 9279
QY 2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLe 2342
Db |||||
9278 -----CAACCCACCCCGGACT----- 9261
QY 2342 uGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluSerProPr 2362
Db |||||
9260 -----CCACC 9256
QY 2362 oLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPr 2382
Db ::::::|
9255 AACCGCGGCCCATACTCTGACCA---TCACCCCGCATACACCCGTCGCGGAACCC 9199
QY 2382 oIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLy 2402
Db |||||
9198 CGAAACCCCAACGGAT---CAATCCCGGACGCTCCCAACGCTCC----- 9156
QY 2402 sAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAl 2422
Db ::::::|
9155 -----ACGACACTCAAAACAAACGCTGCTGCGGATCATCGCCACC-GCCTC 9107
QY 2422 aSerGlyAspArgPro----- 2427
Db |||||
9106 ACGCGGGGACACACCAAAACATCCGTCATCGACATCCCGGTCATACACAAACCCCC 9047
QY 2428 -----ProSerValSe 2431
Db |||||
9046 CTCGCGCACATACGACTTCCCGACACACCCCGGACCGGATCAAAACCCCTCCACATC 8987
QY 2431 rSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpG1 2451
Db |||||
8986 CCAACCCCGATCCACCGGAACCCCGACCGGATCCCGACCCCGCCCAAC---ACCAATC 8930
QY 2451 uAspArgProSerSerAlaGlySerThrProPheProTyr-AsnProLeuIleMetArgL 2471
Db ::::::|
8929 CCACAAACCTTCGCGCACACACCCACCCCGGCAACGAAACGACACCA----- 8883
QY 2471 euGlnAlaGlyValMetAlaSerProProPro-----G 2483
Db ::::::|
8882 -----TCCCGACACACCCACCGGATCATCCACAGCAGCAGCGGA 8843
QY 2483 lLeuProAlaGlySerGlyProLeuAlaGlyProHis----- 2495
Db |||||
8842 GGTGACCTCGCGGACGTTCTCCCGGTTGCGGCGCTGAATTTCTCTGACGTTCCGGGC 8783
```

QY 2496 -----HisAlaTrpAsp 2499
Db 8782 AGCGGCATCGGGGTGGAT 8763

RESULT 27

AAF67220
ID AAF67220 standard; cDNA; 427 BP.
XX
AC AAF67220;
XX
XX 09-APR-2001 (first entry)
DT
XX
DE Novel human polynucleotide, SEQ ID NO: 2976.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
XX breast cancer; lung cancer; cancer detection; ss.
OS Homo sapiens.
XX
FN WO200102568-A2.
XX
XX 11-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US018374.
XX
PR 02-JUL-1999; 99US-0142310P.
PR 02-JUL-1999; 99US-0142311P.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;
PI Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;
PI Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
XX WPI; 2001-091805/10.

Library of polynucleotides for diagnosing a cancerous state of a
mammalian cell and detecting cancer, particularly of the colon or
prostate, comprises 3351 human polynucleotide sequences.

Claim 9; Page 988; 1046pp; English.

The present sequence is one of 3351 sequences in a library of human
polynucleotides. The library is used to detect differentially expressed
genes correlated with a cancerous state of a mammalian cell and can
detect colon, prostate, breast and lung cancer. The library can be used
to produce probes for detection of mRNA and to produce additional copies
of the polynucleotides. The probes can be used for chromosome mapping of
the polynucleotide and for detection of transcription levels. Ribozymes
or antisense oligonucleotides can be generated. The polynucleotides and
their gene products are used as genetic or biochemical markers (e.g. in
blood or tissues) that will detect the earliest changes along the
carcinogenesis pathway and/or monitor the efficacy of therapies and
preventive interventions. The polynucleotides, polypeptides and
antibodies against them can be used in pharmaceutical compositions to
treat the cancers and proliferative disorders such as neoplasia,
dysplasia and hyperplasia

Sequence 427 BP; 94 A; 143 C; 159 G; 31 T; 0 U; 0 Other;

Alignment Scores:

| | | | |
|------------------------|----------|---------------|-----|
| Pred. No.: | 1,03e-13 | Length: | 427 |
| Score: | 689.00 | Matches: | 135 |
| Percent Similarity: | 97.83% | Conservative: | 0 |
| Best Local Similarity: | 97.83% | Mismatches: | 3 |
| Query Match: | 5.21% | Indels: | 0 |
| DB: | 5 | Gaps: | 0 |

US-09-522-753-5 (1-2517) x AAF67220 (1-427)

QY 792 AlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProAla 811

Db 11403 CCACCATCATACGGCGCAGCAGCCACCCACCCACCGCGGCACATCCGGATCCACCACC 11344
Qy 1568 -----ArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeu 1586
Db 11343 GCACCGACTCCAAAC-----CCACCGCCCAACACCGCGATGCGCCGAACTCC 11293
Qy 1586 hrSerThrProArgGluAlaLysSerProHisSerThrValProGluHisHisPro- 1605
Db 11292 ACAAACCCGAAACCATCCCAACAGCGCGCCACCATCATCCGAAACAAACACCCCG 11233
Qy 1606 -----HisProIleSerProTyrGluH 1613
Db 11232 TCACGCAGATTCGGATACCAATACCCACCATCCAAACCCACCCATCAACACACACCC 11173
Qy 1613 isLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheA 1633
Db 11172 ACCACCGACGA-----ATAAACGGCACCACACCCCGCCCGGACAC 11131
Qy 1633 spProThrSerIleProArgGlyIleProLeuAspAlaAlaIleTyrTyrLeuProA 1653
Db 11130 AAACACCCCAACACCGCGACGATCCCTCCACCGCCTCCACATGCGCGAATCGAC 11071
Qy 1653 rgHisLeuAlaProAsnProThrTyrProHisLeuTyrPro----- 1666
Db 11070 GCATAATCCACCGCCCGCGCCGACCCACACACCCATCCCGCTCACACTCACCACAAAC 11011
Qy 1667 -----ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArg 1683
Db 11010 CCCACGACGATCCACATCCCGCGACACACCGCTCAACGACGACCATTCACCGCGCC 10951
Qy 1683 lnThr-----IleIleAsnAspTyrIleThrSerGlnG 1694
Db 10950 ACACCCACCGACACCCACCGCCACCGTTCCTCGGATCCACGATCGGAGTCCG 10891
Qy 1694 lnMetHisHisAsnThrAlaThr-AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db 10890 ACCGACCATTCGACACCGCCACCGCCACACTCCACCATTCACCGCGCGCAACGCC 10831
Qy 1714 SerProArgGluSerSerLeuAlaLeuAsn-TyrAlaAlaGlyProArgGlyIleAs 1733
Db 10830 ACCACCGCGCCCATAGCCACACTCAACACACACCCCGCCACCG----- 10785
Qy 1733 pleuSerGlnValProHisLeuProValLeuValProProThr-----Pr 1748
Db 10784 -----CCGACGAAACCTCAC-----CTCGGAATGACCCACCGCATCCGGATCCGACACACCC 10732
Qy 1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnPro-- 1767
Db 10731 ACCGACATCCCAACCGCG-----CCACGACACCATCACCGCC 10693
Qy 1768 -----PheSerSerArgHisSerSerProLeuSerProGlyGlyProThrHisLe 1785
Db 10692 CACGACACCGGTGACCACTCAACACCGACCC-----ACACCCCTCACCAGCCCGCC 10636
Qy 1785 uthrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAs 1805
Db 10635 CGCACCATCTCCAAACATCCCAATCCCAAC----- 10605
Qy 1805 pArgAspArgGluArgGlyLysIleLeuThrSerThrThrThrValGluHisAlaPr 1825
Db 10604 -----ACCCCGACACCGCCCGACCATCTCC 10579
Qy 1825 otleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerGlyGlyG1 1845
Db 10578 A-----CCATCGACTCAGCAAC-ACCTCGACTGTCGACGAGCGCGGACCAT 10529
Qy 1845 yGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis-SerProIleSerP 1865
Db 10528 ACCACCCACTGCGCAACCTTG-ACCGGAAACACAAACAC-----CACCGACACCAAC 10473
Qy 1865 roArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysG 1885

Db 10472 CAGCAACACCGCAGCCACCGAACCC----- 10447
Qy 1885 lyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSers 1905
Db 10446 -----ACACACACCGAACCCGAA-----GCCACCGCCCGGAAACACCGCATCCA 10401
Qy 1905 exProValArgProAlaAlaThrPheProAlaThrHisCysProLeuGlyGlyThrL 1925
Db 10400 AACCGCGCCACCATCTCGGACACCCCGCCACCAACACG-----ACCG 10359
Qy 1925 eu-AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArg 1944
Db 10358 CCGCATGCTCGAACACCGACCGCGACGACACCAACCGCAACCAACATTCACCAACCGG 10299
Qy 1945 ValAlaArg-----ProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLys 1962
Db 10298 CACCATCGGAAAGACCGCGCACACTCAGCA-----ACCGCGCCGATAAAGCGCAAG 10242
Qy 1963 ProProAlaArgSerGlyLeuGluProAlaSerProSerLysGlySerGluProArg 1982
Db 10241 CCGCGCTCGAAC-----GACCGGAAACACCAACCGGAAACCGCATAGACCCCTAC 10191
Qy 1983 Pro-----LeuValProPro 1987
Db 10190 CCACACCGCATCGGTGTCGGTAAACCGGATCGTGAATTCGGGGGCTGCT 10131
Qy 1988 ValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAla 2007
Db 10130 CMAAATCAGATGCGATTCGTCGCGACACACCCCAACCGCGACACACCGCACGCA 10071
Qy 2008 SerProAspProProAlaProProAla-----SerAlaSerAspProHisArgGluLys 2025
Db 10070 CACGACCCACACCGCGCAATCCCGCGCTCGTCAACCAACTCCACCCAC-----CGAAG 10014
Qy 2026 ThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly 2045
Db 10013 ACCATCCACCAACGGCGACGCA-----CATCCA 9984
Qy 2046 SerSerTyrSerProGluGlyValGluProValSerProValSerProSerLeuThr 2065
Db 9983 CATGCAACGTGCGCGCACACCCCGCGCGCAACGCCA-----TCACCATCTTGATCA 9930
Qy 2066 HisAspLysGlyLeuProLysHisLeuGlu-LeuLeu----- 2077
Db 9929 CAC-----CGGCACACCGCGCAGCGCGCTGCGTATGACCAATATTCGACTTCA 9882
Qy 2078 -----AspLysSerHisLeuGluGlyGluLeuArgProLysGln----- 2090
Db 9881 CCGAACCCAAACACCGCCCGCACACGACCGCGCGCATACGACCCCAACACGCT 9822
Qy 2091 -----ProGlyProValLysLeuGlyGlyGluAlaAl 2101
Db 9821 CAGCTCAATCGGATCACCAACCGAGTCCCGCTCCCATCGGCTC-----CA 9774
Qy 2101 aHisLeuProHisLeuArg-----ProLeuProGluSerGlnProSe 2115
Db 9773 CCACAT-CCACATCCGACACACCAACCGCGCGCCACCAACCGCGCACCAATCACCGA 9715
Qy 2115 rSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLe 2135
Db 9714 CGTGGCGCA-----CCCCGACGCGCG----- 9690
Qy 2135 uAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLe 2155
Db 9689 -----TCMAACCATTCGACCCACCATCTCTGATTCACC 9658
Qy 2155 uSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAs 2175
Db 9657 GCAGAACCCCGCACACCGCGCAAC-----CCGATGACCAACCGCTCCCGCG----- 9612
Qy 2175 pLeuArgArgProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaAr 2195
Db 9611 -----CCACCGACACCGCTCCCAACACCAACCAACCGCGACCTCCCGCCACCC 9562

1029 ProAlaAspLysGluAlaPheAla-----AlaGluAlaGln 1040
13301 CCAGAACTTCGCAACGTCGACGGGTGGTGGTTCGAAGACACCGTCGAGGCAGAC 13242
1041 Lys----- 1041
13241 TCAGACGGTGTCTGTGAGTGGTTCGCAGATTCGAGCGCGGTGAGCGAGTCGAAGC 13182
1042 -----LeuProGly-----AspProProCysTrpThrSerGlyLeu 1053
13181 CGGGTCTTGAAGCCCGGGTGGCGCGCAGCCGATCCACCGTCTCGTGCTCAGGACCG 13122
1054 ProPheProValProProArgGluValIleLysAlaSerPro-HisAlaProAspProse 1073
13121 CGCGCGCGCGCGCC-----GCACCGTTTCGAGGAGGACCGCT 13083
1073 rAlaPheSerTyraProProGlyHisProLeu-----ProLeuG1 1087
13082 GCGGTTCCGGTTCGGGATCCCGCCACCGCTCGCAAGCGCGGCTCGGGGACTCGG 13023
1087 yLeuHisAspThrAlaArgProValLeuProArgProThrIleSerAsnPro----- 1105
13022 TGCCAC-----TTTCCGGCATCGAGCGGCTCTGGGCTCGGGCACCCTCGTCA 12972
1106 -----ProProLeuIleSerAlaLys----- 1113
12971 AGAGCGCGCGCGAGTTCGCAAGACGCGGAGCGAACCCTCCAGTCGACGTCGG 12912
1114 -----HisProSerValLeuGlu----- 1119
12911 CGAGCGTGACAGACATCTATCGGTCGAGCGCTCTGCAGAGCGCGATCGCAGCT 12852
1120 -ArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrsSerG1 1139
12851 CCGGTCCAAAGCGCGTATCCAGCGCGGTTCAGGTTCTGCTCGCGATCCCTTCGCCA 12792
1139 uHisAlaLysAlaProValGlyPro----- 1147
12791 TGCCACCGCGCGCCACAGTCCCAAGCTATGAGGTGGCGGAAGACCTCGCGCAGAC 12732
1148 -----ValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerG1 1166
12731 GTGCTCCGCGCAGCGCT-CCAGCGATCGCTTCGCCCGCGCTACGCGCT-----GG 12679
1166 yValLysGlnGluLeu-----SerProArgGlyGlnAlaGlyProProGlucose 1183
12678 CGCGCGTTCGCCAGACTCGGTAAACCGATGAAGAGGACGAAGCGCTCCAGTCTCTGG 12619
1183 rLeuGly---ValProThrAla-GlnGluAlaSerValLeuArgGlyThrAlaLeuGlys 1202
12618 TGCTGGTGAGTTCGTGAGCAGCAGCGCGCAGCGGTCTTCGACGAGAGACCTCTGTCG 12559
1202 erValProGlyGly-SerIleThrLysGlyIleProSerThrArgValPro----- 1218
12558 ATGTGTCGCGCGCGAGCGTGTGATC---ACACCGTGTGTCAGGACTCCGGCGGTGTGG 12502
1219 -----SerAspSerAla-----IleThr 1224
12501 AAGACCGCATGGGCGGATGCGCGCCAGACGTCGGACAGCGCAGCGCGGTCCGCTTACA 12442
1225 TyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrlsGlyThrIleThr 1244
12441 TCGCAGCGCGCATGTGCTCCTCGACACCGAGC---CGCGAAGATCAGCTTCCAGTTCT 12385
1245 ArgIleIleGlyLysAspSerProSerArgLeuAspArgly----- 1258
12384 GCGCGCGCGCGGCTCGGCACCCCGACGACTGTGTGAGACGAGGTGTCGCGACCGTTG 12325
1259 ArgGluAsp-SerLeuProLysGlyHisValIleTyrlsGlyLysGlyHisValle 1278
12324 CGAGCGGACCGACCGCGGCCACCGCGCCACCGAGACTTCCCGTACCTCCGTCACCGAGTAT 12265

1278 uSerTyrlsGlyGlyMetSerValThrGlnCys-----SerLysGluAspG1 1294
12264 TTCCCGAGGGGGCCATTCGCGCGCGCAGTCCGCGCGCGCGCGTCAACCGACGCCG 12205
1294 yArgSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrlsAspMetMe 1314
12204 TAGATCCGAGCGCGGTACCGCA-----TCTGTCTCTCGCGCTCCGCTC----- 12159
1314 tGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAl 1334
12158 -----CGCGAGGCGGT-----TGGCGAGC 12139
1334 alleProProGluArgHisSerProHisLysLysGluGlnHisIleArgGlyse 1354
12138 TGTCCAGCATCCGGGCAATTCG-----GTACGG----- 12111
1354 rIleThrGlnGlyIleProArgSerTyrlsValGluAlaGlnGluAspTyrlsLeuArgG1 1374
12110 -----ACGGAAGGT----- 12102
1374 uAlaLysLeuLeuLysArgGluGlyThrProProProPro-----ProPr 1389
12101 -----CGACCGCGCGCGCGCGGTCCGGCAGTTCAGGCGGCGCAC 12058
1389 oSerArgAspLeuThrGluAlaTyrlsThrGlnAlaLeuGlyProLeuLysLeuLysPr 1409
12057 CGCCGAAACGCGCACACCTCGGCACGACACAT-----CC 12022
1409 oAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlu----- 1427
12021 GCGCGCTCTCGGGGACACAGCGCGCGCGCGCGTCTCGTGCAGCTCCACAAAGAGTCTCG 11962
1428 -IleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLys 1447
11961 GCTCCCGCGTCCCGATGG---CCTGCACCGCGCGCGCGCGAGCAGCAGCGCGCATTCG 11905
1447 sGluGlySerIleThrGlnGlyThrProLeu-LysTyrlsAspThrGlyAlaSerThrThrG 1467
11904 ATGAGGGGTGTGAGGATCGCGTCTCTGCACAAAGCCGAG----- 11864
1467 lySerLysLysHisAspValArg-----SerLeuI 1477
11863 -----AAGAGACACGCGCGCGCGCAGCGGACCGCGCGCGCGCGTCTGTGTGGGAGCAGT 11809
1477 leglySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAla 1497
11808 TCGCGCATCCGGCGCGCGCGCGCGCGCGGTTCACGCGCGCGCGCGTCTCGCGCGCG 11749
1497 rGluAlaLeuArgAlaCysTyrlsGluLysSerLeuLysSerArgProGly----- 1513
11748 CTGCGAATGAGCGCTGCTGT---GATGTCGCGCAAGCTCCGAGCGCGCGGTCCCGCGG 11692
1513 ----- 1513
11691 TCGCGCATGAGTACGCGCGTGTGTGAGGGGTTCGGGGGTTCGCTCGCGCAACGGTTTC 11632
1514 -----ThrAlaSerSerGlyGlySerIleAlaArgGly- 1525
11631 CAGGTGATCGGATACCGCAGCGCGCGCGCGCGGTTCGGGTGTGTGTGTGTGTGTGTGTGT 11572
1526 -----AlaProValIleValPro----- 1531
11571 GGGT 11512
1532 -----GluLeuGlyLysProArgGlnSerProLeuThrTyrlsGluAspHisGlyAlaProp 1550
11511 GTCGGAGGTTCACCAACCG 11461
1550 heAlaGlyHisLeuProArgGlySerProValThrMetArg-----GluProThrPro- 1567
11460 TCAACTCAT-----CCCGCGGAGGAGCGCTTACCACCGCGCGCGCGCGCGCGCGCGCG 11404
1567 ----- 1567

```
Db 15208 ACCGCGCCGCGCGCACCACTCCAAACA...:::ATCCCAATC 15170
Qy 423 TyrIysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGlu 442
Db 15169 CACAACCCCGA...:::CAACGCCGACCACTCCACCAT 15134
Qy 443 LysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThr 462
Db 15133 CGACTCGCGNAC...:::CTCCGACTCGTCAGCAGCGCGCAC 15092
Qy 463 ValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu 482
Db 15091 CATACCCACCCACTG...:::CCAACCTGACCGGAACACAAACACCAACCCAGCCGAC 15038
Qy 483 Val...:::ArgArgSerTyrArgArgArgGlyLysSerGlnGln 496
Db 15037 ACCACCAACACCCGACCGCACCGAACCCACCAACCAACCCAGCCGAC 14979
Qy 497 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 513
Db 14978 CAGAACCGACCGACCAAGACCAACCAACCC-TCTTCCACACCCACCCGAAACACGACC 14920
Qy 514 ---ArgSerSerGlnGlnGluLysAspGluLysGluLysGluLysGluLysGlu 532
Db 14919 GCCGATGCTCAACACCGACCGCGACGACCAACCAACCAACCAACCAACCA 14860
Qy 533 GluGluLysProGluValGluAsnAspLysGluAspLeuLysGluLysThrAsp 552
Db 14859 GAACCAACCGGAG...:::CCACCAACCGTCAGAC 14830
Qy 553 ThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAla 572
Db 14829 AACCCCGGCACACTCAGCAACCGACCGCATACGCC...:::CGAAAGCCGCC 14779
Qy 573 AsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
Db 14778 ACCGAACGACCGACACCCACCGGAACCAACCGACGACTCTCTGCCAACACCACTCG 14719
Qy 593 GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer 612
Db 14718 CCGCATCCACGACACCGGCATCGGAATCGGAA...::: 14686
Qy 613 SerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg 632
Db 14685 ---TCGGAATCGGAATCGGAATCGGCA...::: 14662
Qy 633 AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsn--- 651
Db 14661 ---GGATCGTGAACCTCGGGGCGCTGCTCCAAAATC 14629
Qy 652 ---PheTyrPheAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGln 668
Db 14628 AGATCGCATTTCTGTC-CGACCCCAACCGCGACACACCCCGACGACGACGAC 14570
Qy 669 HisLysLeuLysMetGluLysGluArgAsn...::: 678
Db 14569 CACCACGGCCATTCGCGGCTCCGTCACCACTCCACCCACCCCGAAGACCAATCCAC 14510
Qy 679 ---AlaArgArgLysLysLysLysAlaProAlaAlaAla 690
Db 14509 CAACGGCGACGCACATCCACATGCAACGTTCGCGGA-CACCCCGCGCGCAACGCCA 14451
Qy 691 SerGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGlyVal 710
Db 14450 TCACCATTTGATCACCACCGCCACACCCGCGCGCTGCTGATGACCAATATTCGACT 14391
Qy 711 SerGlyAsnGluGluMetValGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 14390 TCACCGAACCAACACACGCCCCACCCACGACCCCGCCCATACGAC--- 14340
Qy 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp-----Thr 748
:::|||||
```

```
Db 14339 -----CCAACAACGCTCAGCTCAATGGATCAACCAACC 14304
Qy 749 GluSer-----IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn 765
Db 14303 GAGTCCCGCTCCCATGCGCTCCACCATCCACAT-----CGACACACCCAAAC 14253
Qy 766 GlyProLysProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro 785
Db 14252 CCGCGCGCACCAACCGCGCACCAATCACCGGACGCTCGGCA-----CCCCGACGCG 14199
Qy 786 ProArgArgThrSerArg-----AlaProIleGluPro 796
Db 14198 CCGTCAAAACCATTCGACCCACCATCTGATTACCGGAGAACCCCGCACCA---CCGCA 14142
Qy 797 ThrProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSerProSer 816
Db 14141 ACACCCGATACCAACGCTCCGCGCACCAACCAACCTCCAAACACCAACCCCGCAC 14082
Qy 817 AlaProPro---ProValValProLysGluGluLysGluGluThrAlaAlaPro 835
Db 14081 CTTCCCGCCCGCTCCCATCGCACCCGCAACCGCGCAATCACCGTCAACC 14022
Qy 836 ProValGluGluGluGlu-----GlnLysProProAlaAlaGlu----- 850
Db 14021 ACAACCCACCTGCGGAGAAACTCCAAACAAACCCCGCGCAATCACCGTCAACC 13962
Qy 851 -----LeuAlaValAspThrGlyLysAlaGlu 859
Db 13961 CACCCGCAACGCGCACATCATCTACCCAAACCGCAACCTCGACACGCAAAATGCAACG 13902
Qy 860 GluProValLysSerGluCysThrGluGlu----- 869
Db 13901 CCACCAACGACGACGACGCGCTATCAACCGAAATCGAGGACCTCCAAACCCCAACA 13842
Qy 870 ---AlaGluGluGlyProAlaLysGlyLysAsp-----AlaGluAlaAlaGlu 884
Db 13841 CATACGAAACACGACCGCGACGCGCTCACCGACTTGCCTGAGCAGATGAGCTCTGA 13782
Qy 885 AlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGlySerGlyArgAlaThr 904
Db 13781 GTTCTCGGAGGCTCTGTGAGACGCGGCGGCTACTCTCTGTGGGTGAGCGCGGATACA 13722
Qy 905 ThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAsp 924
Db 13721 ---CGCGGCTCGGCTACCCGCGAC-----CCGCCGAT 13689
Qy 925 GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeu 944
Db 13688 CGATTCCGCGACGCT-----CCAAACGCTCCACG 13659
Qy 945 LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys 964
Db 13658 ACGTCTCAACAAACGCGTGTGTGGATTCATCGCCACCGCTCAACGCGGCGACAC 13599
Qy 965 Gln-----LeuLysGlnArgAlaAlaIleProIleGlnVal 978
Db 13598 CAAGAAGCTCGCATCGAACATCCCGCGTCATACAAACCCCTCCCGGCGATACG 13539
Qy 979 ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro----- 996
Db 13538 ACTTCCCGCGCGCGCGCGGATCAAGAGGAGTCCAGTCCAGCCCGCGTCT 13479
Qy 997 -----ProAlaProProProGlnAsn----- 1004
Db 13478 CGGGGGCTGGCGCATGACGTACGACCGCGGACGAGGAGTCCAGAGTTCTCTGGGAG 13419
Qy 1005 -----LeuGlnProGluSerAspAlaPro----- 1012
Db 13418 TGTCTCGCTCGCGGGGTACGCGGCGCATCGCGAGATGGCAATAGGTCTCTCGACG 13359
Qy 1013 -----GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPro 1028
Db 13358 GCACGGCTGCGGAGTCAACCGGA---CTTCTCGGAGAAAGCGCGGACGAGGCGCCT 13302
```



```
Db 3860 GTCCATCCGACACCGCGAAACCGCGACACGCCCATCATCCCGACACCCACCTGC 3801
      |||||
Qy 2203 YLysArgSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrPr 2243
      :|||:
Db 3800 CGAGAAACTCCACAAACCCCGCG- 3773
Qy 2223 uProValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrPr 2243
      ||| |||||
Db 3772 -CCGCAATACCGGTACCCCGACCGCGCAACCGCATCATCACTACCCACCGCAACCC 3714
      ||| |||||
Qy 2243 oLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProG1 2263
      ||| |||||
Db 3713 T-----GACAGCGCAATGACCGCCACCGCACGACGACGACGCGGTATCAACC 3663
      :|||:
Qy 2263 yAsnThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
      :|||:
Db 3662 GAAATCGAGCGACCTTCAACC----- 3641
Qy 2283 lLysSerLysLysGlnGluLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluPro-- 2302
      |||||
Db 3640 -----CCAAACATACGAAACACGACCGCGAC 3615
Qy 2303 -GluTyrAsnIleSerGlnProGlyThrGluLeuPheAsnMetProAlaIleThrGlyTh 2322
      :|||:
Db 3614 ACCACATCGCGGAGTCCCGTCAACA-----ATAACCT- 3578
Qy 2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLe 2342
      ||| |||||
Db 3577 -----CAAAACCCACCGCGACT----- 3560
Qy 2342 uGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProPr 2362
      |||||
Db 3559 -----CCACC 3555
Qy 2362 oLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPr 2382
      :|||:
Db 3554 AACCGCGCCCATCTCTGACCCA---TCACCCCATATACACACCGTCCGCGAACCC 3498
Qy 2382 oIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLys 2402
      :|||:
Db 3497 CGCAAAACCAACGGAT---CAATCCCGCAGCGCTCCACGCGCTCCC----- 3455
Qy 2402 sAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAl 2422
      :|||:
Db 3454 -----ACGACCTCTCAAAACAAACGCTGCTCGGATCATCGCCACC-GCCTC 3406
Qy 2422 aSerGlyAspArgPro----- 2427
Db 3405 ACGCGCGCACACACCAAAACTCCGATCGAATCCCGCGTCATACACAAACCCGCC 3346
Qy 2428 -----ProSerValse 2431
Db 3345 CTCGCGCATACGACTTCCCGACACACCGCGACCGGATCAAAACACCTCCACATC 3286
Qy 2431 rSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrG1 2451
Db 3285 CCAACCCCGATCCACGGAAACCCCGACACCGCATCCCGACCCCGCAAC---ACCAATC 3229
Qy 2451 uAspArgProSerSerAlaGlySerThrProPheProTyr-AsnProLeuIleMetArgL 2471
      :|||:
Db 3228 CCACAAACCTCCGACACACACCGCCCGGAAACGACACCCA----- 3182
Qy 2471 euGlnAlaGlyValMetAlaSerProProProPro-----G 2483
Db 3181 -----TCCCGACACACCGACCGGATCATCCACGACGACGCGGA 3142
Qy 2483 lLeuProAlaGlySerGlyProLeuAlaGlyProHis----- 2495
Db 3141 GGTGACCTCGCGAGGTCTCGGTGCGGCGCTGAATTTCTTCTGCAGTTCGCGGC 3082
Qy 2496 -----HisAlaTrpAep 2499
      |||||
```

```
Db 3081 AGCGGCCATCGGGTGGGAT 3062
RESULT 26
AAD55810/c
ID AAD55810 standard; DNA; 60196 BP.
XX
AC AAD55810;
XX
DT 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
XX Micromonospora carbonacea polyketide synthase (PKS) type I gene cluster.
XX Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
XX gene; ds.
XX Micromonospora carbonacea.
XX Key Location/Qualifiers
XX CDS 1..1683
FT /tag= a
FT /product= "Polyketide synthase #1"
FT /note= "CDS does not include start codon"
FT /partial
FT CDS complement (1728..2522)
FT /tag= b
FT /product= "Polyketide synthase #2"
FT /complement (2629..3861)
FT /tag= c
FT /product= "Polyketide synthase #3"
FT /tag= d
FT /product= "Polyketide synthase #4"
FT /tag= e
FT /product= "Polyketide synthase #5"
FT /tag= f
FT /product= "Polyketide synthase #6"
FT /note= "CDS does not include start codon"
FT /partial
FT CDS 24993..36230
FT /tag= g
FT /product= "Polyketide synthase #7"
FT /tag= h
FT /product= "Polyketide synthase #8"
FT /tag= i
FT /product= "Polyketide synthase #9"
FT /tag= j
FT /product= "Polyketide synthase #10"
FT /note= "CDS does not include start codon"
FT /partial
FT CDS 47794..49083
FT /tag= k
FT /product= "Polyketide synthase #11"
FT /tag= l
FT /product= "Polyketide synthase #12"
FT /tag= m
FT /product= "Polyketide synthase #13"
FT /transl_except= (pos:51221..51223, aa:Xaa)
FT /note= "Xaa corresponds to amino acids from position 452-1811"
FT CDS 51506..53416
FT /tag= n
FT /product= "Polyketide synthase #14"
FT /note= "CDS does not include start codon"
FT /partial
FT CDS complement (54569..53358)
```


Db 8573 TGCCCTTGC-----TCAGCGCGCTTCTCTCGCTGAGGCGGTGAGC 8532
Qy 1911 AlaThrPhePro-ProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyPr 1930
Db 8531 GCATGAGCGCGCAGCTCGCTCTCTCTCTCAGCGCGCTT-----GCTCTCTC 8478
Qy 1930 oThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluAr 1950
Db 8477 AGCCAGCTCGAGCTCGCTCGTCCACGG-----CCGCGATGTGCAGATCACCTCG 8424
Qy 1950 gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProPro-----1964
Db 8423 GCCCGCGCGCGCGAGCGCACAC-----GCACACCTCCGGAAGCTTCCGGCAAGC 8370
Qy 1965 -----AlaArgSerGlyLeuGluPro-----1971
Db 8369 GGAGGTGACACACGAGCGAGGTGCCCCCTTATATCCGGACTACCCGTCACACGCTCGGA 8310
Qy 1972 -----AlaSerSerProSerLysGlySerGluProArgProLeuValProValSe 1989
Db 8309 CGGTTTGGCAGCGTCTCTCATACAGCCAGCTCGCGCGCGCTTCTGCTGCTTCTC 8250
Qy 1989 xGlyHisAlaThr-----IleAlaArgThrPr 1998
Db 8249 TGGC-----GCCACGCGCGAGGTACGGCAGCGCGCGCGCGCTTCCGCTGCTCA 8193
Qy 1998 oAlaLysAsnLeuAlaProHisHisAlaSerProAspProAlaProAlaSerAl 2018
Db 8192 GGAGCTCCAGAGCTCCGCAACGCCATCTCCCTCGCCCTCCAGCGCGATTCACAGC 8133
Qy 2018 a-----SerAspProHisArgGluLysThrGlnSerLysProPheSe 2032
Db 8132 GCCTGCTCGCGCTGGAGACATTACCTCGTGGCGCGCTTTCGGCGCTTCAGCCAG 8073
Qy 2032 rIleGlnGluLeuLysArgSer-----LeuGlyTyHisGlySerSe 2047
Db 8072 TAGCGTGGCGGTGGAAAGCGTACGTGGCAGCTCCACGGCGCGCGCGCGCGCGAC 8013
Qy 2047 rTySerProGluGlyValGluProValSerProValSer-----SerProSerLeuTh 2065
Db 8012 A-----GCACCGCGCGCGAGTCCACTCGTGGCCCTTGACAGTGCACGTGCGCCACCGCGTC 7956
Qy 2065 rHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGlyGly 2085
Db 7955 GCCAGCGCGCGCTTCTCTCTGCTCTC-----GCCGCGAGGCTCGTCAAAACGTGCGCTCG 7899
Qy 2085 uLeuArgProLys--GlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuPr 2104
Db 7898 GCTCCTTCCGGCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7839
Qy 2104 oHisLeu-----ArgProLeuProGlu-----SerGlnProSerSerProLe 2119
Db 7838 ACGTATGTGCTCACCCCGCGCGCGAGGTGGCATCCGCTCCAGGAAGCGCACCG-- 7781
Qy 2119 uLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIl 2139
Db 7780 -----CCTCGCGCGCACTGCTCACCAGTACTCGCGCGCACATCAGCGCTTCTCGCCAC-- 7727
Qy 2139 eSerGluValIleThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaPro-- 2158
Db 7726 -----CGAGTTCG 7695
Qy 2159 -----LeuProAlaProLeuTyThrSerPheProGlyAlaSer-- 2170
Db 7694 AGCGTGGCGCGCGGTACGTGCACTCCCGCGCGCACTTCCCGAACTCTCCAGCATCCCG 7635
Qy 2171 ----CysProValLeuAspLeuArgArgProProSerAspLeuTyThrLeu-----ProPr 2187
Db 7634 TCCATGTGCGCGCTGTGGAAAGCGGTGGACACACGACGCGCGCGCGCGCGCGCGCGCTGC 7575
Qy 2187 oProAspHisGlyAlaProAlaArgGly-----SerProHisSerGluGlyGly 2204
Db 7574 GCCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTGCTCCCGCTCAGCACCGTCTGC 7515

RESULT 25

AAD55815/c

ID AAD55815 standard; DNA; 13416 BP.

XX

Qy 2204 sArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGlyGlyGlyGlyGlyGly 2224
Db 7514 ATCGGCGCGTTCAGCCCGCGATGCTCAGTGCCTGCGCCCGACCTCCGACGAGCGC 7455
Qy 2224 oValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyPrProLe 2244
Db 7454 CGTGACCTCCGGCTCCGAGGCTCCACGACCATCGCTCCCGCGCGCTCGCACCT 7395
Qy 2244 uLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAs 2264
Db 7394 TGATCAGCGGACCTCGGGGAC-----CTAGCTTCGCTCGCTCGCAAGG 7347
Qy 2264 nThrSerGlnProPro---AlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
Db 7346 CTGACACACCGCGCACATGCGCAGCGCTCAGCTCTCTATCGAGTGCCTCCGACGCGCA 7287
Qy 2283 lLysSer-lysLysGlnGluIleAsnLysLysLeuAsnThrHis---AsnArgAsnGluP 2302
Db 7286 GCGGGCTTCAGCCCCCAGCACTCCCACTGACGGTACAGCGCCACTCCAGCGCAACAGC 7227
Qy 2302 ro-----G 2303
Db 7226 CCGGGCTGCGGTACTCCGCTCCAGCAACGCTCTCTCTCGCTGCGCGCGCGCGC 7167
Qy 2303 luTyraenLysSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrG 2323
Db 7166 AACACCACTCTCTCAACCCACCGTTCGAGATA-----7135
Qy 2323 lLeuMetThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuG 2343
Db 7134 -----CCTGTCAGCGCTCGCACACCT-CGTGACGCGCGCGCGCGCGCGACAC 7090
Qy 2343 luAlaIleAlaArgLysAlaLeuMetGlyLysTyArgGlnTrpGluGluSerProProL 2363
Db 7089 GGGGTACACTTCGTAAGCC-----TCTTCCCATCCC 7057
Qy 2363 euSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla-----AlaM 2381
Db 7056 GAGCGCTGCTGCTGCC-CGCGCGTGAACAGCACCGCAAGCTTCCCGCTCGCTTCGCGC 6998
Qy 2381 etProile---ThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2400
Db 6997 TCCGCTCACCAACCGCGCATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6957
Qy 2400 lyGlyLysAlaLysValSerGlyArgProSerSerArgLysAla-LysSerPro---Ala 2418
Db 6956 -----GGCGAGAGACCTCCACAGCTCCGCGCGCGCGCGCGCGCGCGCGCGCG 6914
Qy 2419 ProGlyLeuAlaSerGlyAspArgProProSerValSerValHisSerGluGlyAsp 2438
Db 6913 CCG-----6911
Qy 2439 CysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGly 2458
Db 6910 -----ACGCGCGCGACT-----CGAAGTTCGCTCGCGTGCA 6881
Qy 2459 SerThrProPheProTyraenProLeuMetArgLeuGlnAlaGlyValMetAlaSer 2478
Db 6880 GCGCGCGGTGGCACACCGTCCG-----ACCACC 6851
Qy 2479 ProProProGlyLeuProAlaGlySerGlySerGlyProLeuAlaGlyProHisAlaTrp 2498
Db 6850 CCACTTCCCGCTGCTCTTCCAGCGCACTTCG---CCCACCGCGCGCGCTCGCGCATTC 6797
Qy 2499 AspGluGluProLysProLeuLeu 2506
Db 6796 -----CCGCGCGCTCGTCTC 6782


```
Qy 319 AsnProArg-----ArgAlaLysGluSerLysValArgGlu----- 331
Db 14697 CGTCCGAGCGTCCGCGAGCATCGAGGAGAACACGACGAGCGCGGAGAGATCGAGATCCTC 14638
Qy 332 -----TyrTyrGluLysGlnPheProGlu----- 339
Db 14637 TGTCAAGCTGTCAGAGTGCAGCGCCCGGTTCACTTCGGCGCCACGACCCCGGAGAGACG 14578
Qy 340 -----IleArgLysGlnArgGluLeuGlnGluArg 349
Db 14577 CTCGGGCGTCTGCGCGGTGATGACGCCATCGTCGACGACGCGCGCGAGGTGCAGCACCGC 14518
Qy 350 MetGlnSerArgValGly----- 355
Db 14517 GGTCAAGCGATGTCGGGCGCTCGATCGCGGCGCAGCACGCGCGGACCTCTCCCTCCGTTGCA 14458
Qy 356 -----GlnArg----- 357
Db 14457 CACGTCAACGCCGCCACCGTCACGGTCTCGGCACGGAGCTTCTCCAGCGCTGCACGAA 14398
Qy 358 -----GlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIle 375
Db 14397 GCCTGGGGCCCCGGCGCTCCAGCCACGCGCGGAGCGTCAAGCACGAGGTGCGCGACCCC 14338
Qy 376 Ile-----AspGlyLeuSerGlu----- 381
Db 14337 GTGTGGGCGCACAGATGCTCCGGGACGGCTTGACCCAGCTCGCCTGTCTCCGCTCAC 14278
Qy 381 ----- 381
Db 14277 CAGGACGTCGCGCTCGAGCGCGCGCGGTGAGCCCTTCGCGTCCGCGCGG 14218
Qy 382 -----GlnGluAsnLeuGlu-----Lys 387
Db 14217 CACGCGCACGCGCGCGCCATCGCGCGCGCCACGACGCGCAAGCTCCGCGCTCGT 14158
Qy 388 GlnMetArgGlnLeuAlaValIleProProMetLeu-----TyrAsp 401
Db 14157 CGCGTGCACGCGCGCGCGCACGAGCCCGCTCCAGAGGCTCGGTCCCAAGCTCGAT 14098
Qy 402 AlaAspGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
Db 14097 CAGCGCAGCGCGCGCTCGGGGTGCTCGCTCGCGCGGTGCGAAACAGCCGCCACAGCGG 14038
Qy 422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe--- 440
Db 14037 -----CGCGTCCCGGCTTCGACACCGTCCCGGAGCGCGCTGACCGCATCTCG 13984
Qy 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db 13983 CGTCACCCACACAGCTCCACCCCTCGAGCGCGGGCTCCGACAGCCACGCTTGACG--- 13927
Qy 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db 13926 -----CAGCGACAGCGCTGTCGCGT 13906
Qy 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
Db 13905 CGCTCGTGCACGATATCATCATGTCGAGCGGTGTGAGTGTCCGAGCGCATGATGATCGAC 13846
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSer---GlnGluGlu 519
Db 13845 CAGCACCCGCTCAGGCAATCGGGTTCCTGCTCGATCCGCGCGAGCAATGATCATGAGGCC 13786
Qy 520 LysAspGluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGlu 539
Db 13785 TGCATCGCTCCACACCCAGCGGTGCGCCAGTCGCGCTCGCGCTCGCGCGCACGAG 13726
Qy 540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAsp 559
Db 13725 CACCACCGAGGATCGACCTTCGAGCGCGCTCCACGAGGCTCAGAGGTTCAGAGTCCAC 13666
Qy 560 GlulysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgLys 579
Db 13665 CCAGTACAGGTGCTGCGCCTCGCATCGTCCGCGCGAGTCTCTCGCGCGTCCGCGCG 13606
Qy 580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
Db 13605 GCGCAGATGAGGACACCGACGCTCACACCGGTGCGCGCGCGCATCTGTAAACGCGAG 13546
Qy 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyr---ThrGluGlu 618
Db 13545 CGAGCGCGCTGCTGTCGGAGTC-----TGTGGCTGTTGGAGCTCCAGCT 13498
Qy 619 GluMetGluThrAlaLys---LysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIle 637
Db 13497 GAGCGGAGTCTGCTCCCGCGTGTCAAGCGCACGCTCCAGCACGAGAAAGGAG 13438
Qy 638 AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys 657
Db 13437 CAGCACGCTGTC-----CGGCGCTGATACCTC----- 13411
Qy 658 LysArgGlnAsnLeuAspGluIleLeuGlnGln-----HisLysLeuLysMetGlu--- 674
Db 13410 ---CGCAACGCTCGACCATCTGATGAGCGCGCGTCCATCAGCGCGGATGAGACCCC 13354
Qy 675 -----LysGluArgAsn 678
Db 13353 GTAGTCTCGCGCTGTCGTCGCTCTTTGGGCAACACACCGCGCGGAGAGTCCG 13294
Qy 679 AlaArgArgLysLysLysLysAlaPro-----AlaAlaAlaSerGluGluAlaAla 695
Db 13293 GTCTCGACGACAGCTCCAGAGCCCTTGAAGCGCGCGGTAGCGGAGTCCACGCTC 13234
Qy 696 PheProProValValGluAspGluMetGluAlaSerGly----- 709
Db 13233 ATGACGCGCTCTGTAGAACCCGAGAGGTCCACCGCTCGCGCGCGGACCTGSCCAGT 13174
Qy 710 -----ValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeu 724
Db 13173 CGTCAGCTCATCGAGCTCGCGGAGGTGGCGGCTCTCGTCTCGTCAA----- 13126
Qy 725 HisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsn 744
Db 13125 CACCCCGT-TGCGTGTGTCACCCAGGGGCGATCTCCGCTCTCGGCTCGGCTGTGT 13067
Qy 745 SerSerAspThrGluSerIlePro----- 752
Db 13066 ACAGCCCAACGACGCGCGCGCGCGGTGCGCGCGCGCGCGAGCTGCGAGCGCA 13007
Qy 753 -----SerProHisThrGluAlaAlaLysAspThr----- 762
Db 13006 CGGCCACGTCCTCGCGAGCATCAGCGCTCGGAGATCGTGAGCTCCGACAGCGCGCGC 12947
Qy 763 -----GlyGlnAsnGlyProLysPro----- 769
Db 12946 TGCCCGCTGCGCGCGCGCGCGCGAGTCCAGCATCCCGTCCGCGCGGAGAGCA 12887
Qy 770 -----ProAlaThrLeuGly--- 774
Db 12886 CCTCGCCAAACACCGCATGTCGCGAAGCCAGCGGTCTCTCGCCGACGACGCGCGCTG 12827
Qy 775 -----AlaAspGlyPro 778
Db 12826 TGAATAGATGGCGCTCGCGCTGGCCAGTCTGTTGCGCTCCGAGCAGCGAGTCCCGC 12767
Qy 779 ProPro---GlyProProThrProProArg----- 787
Db 12766 ACTCCCTGAAGCCCGCGAGCCACGTCGCTCGGTACCGCGCTTCGCGCTTCAGCGAGTAC 12707
Qy 788 -----ArgThrSerArgAlaProIleGluProThrProAlaSerGluAla 802
Db 12706 GCTGCCGCTGAAGCGGTAGCTCGGACGCTCCAGCGCGCGCGCGCGCGCGCGCGAG 12647
Qy 803 ThrGlyAlaProThrProProAlaProProSerProSerAlaProProValVal 822
Db 12647 GCTGCCGCTGAAGCGGTAGCTCGGACGCTCCAGCGCGCGCGCGCGCGCGCGCGAG 12587
```



```
Db 22349 TTTGACTTCGACGAGCAAGGTACAGCGGTGCTCGGCTTCGCGCCCGAGGCCGCAACACT 22290
QY 1913 - - - - - ||||| ||||| ||||| ||||| ||||| - - - - - PheProPr 1915
Db 22289 TCAGCGAGCGCTCGGCTCGATCGATCTCCGAGGCTGTCGCCGCTACGATCGCGCTCC 22230
QY 1915 oAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluPr 1935
Db 22229 ACCGCTCGATGTCCTCGGCGCAGCGCCGACGACGACGCGCTGCTGATCACCCGC 22170
QY 1935 oValLeuLeuProLysGluAlaProArgValAlaAArgProGluAArgProAlaAspTh 1955
Db 22169 TGCTCGCGCGCGCTTCGCGCGCTGAGGCTTCGCGACCGCTGCTGCTCACCGCT 22110
QY 1955 rGlyHisAlaPheLeuAlaLysProProAlaAArgSerGlyLeuGluProAlaSerPr 1975
Db 22109 GACCCAGCGA - - - - - CCACGCAAGCACACGGTCAACGTCGCGCGCGCTCC 22062
QY 1975 oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl 1995
Db 22061 GACAGCGCTTCAGCACGATCCCGCACCCCTCCGACGACCGCGCGCTCGCGCTGC 22002
QY 1995 aArg - - - - - - - - - - - - - - - - - - - - - - - - - - - ThrProAl 1999
Db 22001 CCAGAGAAGCTCTTCAGCGCGCGCTCTCGGCGCATCCCTTGAGCGCGCTGAACCTCAAG 21942
QY 1999 aLysAsnLeuAlaProHisAlaSerProAspProProAlaPro - - - - - 2014
Db 21941 AACAAACGCGGGTGCTATCATCCGTCACCGCCGCGCGCGCGCGCGCTGCATCAACC 21882
QY 2015 - - - - - - - - - - - ProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLy 2029
Db 21881 TGACGCAACCGCTGACGCGAGTGACGACACACGACGACGACGACGCGCTGCC 21822
QY 2029 sProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyHisGly - - - - - SerSerTy 2048
Db 21821 ACCGTATCGCTGGCGCAT - - - - - GCAGCGCGGACGACGTAGGCGCACACGCGCTGAGAGCAGC 21765
QY 2048 rSerProGlu - - - - - - - - - - - GlyValGluProValSerProValSerProSerLe 2064
Db 21764 CTCGCCCGCTTCGATGCCCCGCTACCGCTCCAGCATCTCCAGCCCGCACCCCTGTAA 21705
QY 2064 uThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGlu 2084
Db 21704 GCGCATAGTCCG - - - - - - - - - - - - - - - - - - - - - - - - - - - 21692
QY 2084 yGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuPr 2104
Db 21691 - - - - - - - - - - - AACCGCGCGCGCCAGATACACCCCGTGGCGCTCCGCTCAGCGCGACGGG 21639
QY 2104 oHisLeuArg - - - - - - - - - - - ProLeuProGluSerGlnProSerSerProLeuLeuGlnTh 2122
Db 21638 CGACGCGCGGACGCTCCATGCT - - - - - - - - - - - - - - - - - - - - - - - - - 21599
QY 2122 rAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluVa 2142
Db 21598 -GCACACCGCTGCTGGGATCATCGCTCGCGCTCTCGGCGC - - - - - 21554
QY 2142 lIleThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPr 2162
Db 21553 - - - - - - - - - - - ATATCCGAGAAAGTCCGCTCGA - - - - - ACAGGTCGATGTCGCCGAAG 21510
QY 2162 oLeuTySerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAs 2182
Db 21509 AATCCACCTCGCGCGCTGCTTGCCT - - - - - - - - - - - - - - - - - - - - - 21479
QY 2182 pLeuTyLeuProProAspHisGlyAlaProAlaAArgGlySerProHisSerGluGl 2202
Db 21478 - - - - - - - - - - - CCGCTTCGCGCTCGGGTCTGAGACCGAAGCGTCTCCAGCGCGCAGGG 21429
QY 2202 yGlyLysArg - - - - - - - - - - - SerProGluProAsnLysThrSerValLeu - - - - - 2215
```

```
QY 1268 allIeTyrgluGlyLysLysGlyHisValLeuSerTyrgluGlyGlyMetSerValThrG 1288
DB 24136 CCGCAAGGTACGCGAGCGCGCGAGCGTCTCG-----CGCA 24098
QY 1288 lncysSerLysGluAspGlyArgSerSerGlyProHisGluThrAlaAlaProL 1308
DB 24097 CGTGTGTCAGGAGCTCCAGCAGATCCGCAACGCGCTTCGCTTCGACCGCGTC-CCAC 24039
QY 1308 ysArgThr-----TyrAspMetMetGluGlyArgValGlyArgAlaIles 1323
DB 24038 AAGCAGCTCCGGCCACGAGACATTCGCTGTGCGCGCGCTGCTTGGCGC----- 23986
QY 1323 erSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg----- 1339
DB 23985 -----CTCAGCCAGTAGCGTCCGCTGCGTGAAGCGGTACGTCCGC 23946
QY 1340 -----HisSerProHisHisLeuLysGluGlnHis----- 1349
DB 23945 AGCTCACGGCGCGCGCGCCAGCAGCAGCAGTGGGCGCCAGTCCACCTCGTCCGCC 23886
QY 1350 -----HisIleArgGlySerIleThrGlnGlyIleProArgSerTyrgluAlaG 1367
DB 23885 TGCAGTGTCACTGTGCGCACCGCGCTGCGAAGCGCGCTTCTCTCTCTCTCTGCGCGC 23826
QY 1367 lncIuAspTyrgluGlyArgGluAlaLysLeuLysArgGluGlyThrProPro----- 1385
DB 23825 AGCTCGTCAACAACGTGCTGCTCGCTCCCTC-----CGCA-GGCACC----- 23773
QY 1386 -----ProProProPro-Ser 1390
DB 23772 CGCGCACAGCAGCCATCCCGCCCGCACACTCGAGTATGTCTCACCCCGCGCGCGAAG 23713
QY 1391 Arg-----AspLeuThrGluAlaTyrgluThrGlnAlaLeuGlyProLeuLysLeuLys 1408
DB 23712 CGTGGCATCCCGTTCAGGAAGCGCACCGCTCGCGCACCTG---CCTCACCCAGTACTC 23656
QY 1409 ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIle 1428
DB 23655 GGCGGACATCAGCGCTTCTTGGCCACCGAGCTCGCGCTCACCGCGCTCACCGCGCAG 23596
QY 1429 Pro-----ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArg 1444
DB 23595 CCGTGGCGCGCGTACGTGCACCGCGCGCGCACCTTCCGAACT-CCTCCAGCATCCGCT 23537
QY 1445 ProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrgluAlaSer 1464
DB 23536 CCATGTGCGCGTGTGA---ACGCGTGCAGCA-----CACGCGAGCGCGCG 23492
QY 1465 ThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPhe 1484
DB 23491 TGGCGCGCGCTCGCGCTCCAGCGCTC-----GGCCACCGCGAGCA--- 23450
QY 1485 ProProValHisProLeuAspVal---MetAlaAspAlaArgAlaLeuGluArgAlaCys 1503
DB 23449 CCGCGCTTCGCTCCCGCTCAGCACCGCTCTGCATCGCGCGGTTCAGCGCGCGATGTCA 23390
QY 1504 TyrGluSerLeuLysSerArgProGlyThrAlaSerSerSerSerGlySerIleAla 1523
DB 23389 GTGCGCCCTCG---CCCCGACCTCCGACAGCGCGCGCT-----GCA 23351
QY 1524 ArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyr 1543
DB 23350 CATCCGCTCCGAGGCTCCACCGACACCATCGCTCCCGCGCTCGCACCCCTCGCATCA 23291
QY 1544 GluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArg 1563
DB 23290 GCGACCGCGCGCGCACACTAGCTTCTGCTGCTCGTCCGCAAGGTCA----- 23246
QY 1564 GluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArg 1583
DB 23245 GCACACCGCGCACGTGCGCAGCGCTCAGCTCTCTATCGAGTGCCTCCCGCAGCGCGCGG 23186
```

```
QY 1584 LysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHis 1603
DB 23185 -----GCTTACGCCCC-----ACGACTCCCACTGAC 23159
QY 1604 HisProHisProIleSerProTyrgluHisLeuLeuArgGlyValSerGlyValAspLeu 1623
DB 23158 GTTACAGCCCACTT----- 23144
QY 1624 TyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeu 1643
DB 23143 -----CCAGCGCAACACGCCCGGTGGTGTACT 23114
QY 1644 AspAlaAlaAlaAlaTyrgluProArgHisLeuAlaProAsnProThrTyrgluHis 1663
DB 23113 CCGTCCGCTCCAGCAACGCTCTCTCTGCG-----TGCGCGCGCGCGCA----- 23069
QY 1664 LeuTyrgluProTyrgluLeuArgGlyTyrgluProAspThrAlaAlaLeuGluAsnArgGln 1683
DB 23068 ---ACACCACTCTCTCAACCCAGCGTGCAGATGCGCGTGCAGCGCTGCG----- 23021
QY 1684 ThrIleLeuAsnAspTyrgluThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMet 1703
DB 23020 -----ACACTGTGCAAGCGCGCGCAACACCGGGTACTACTT 22982
QY 1704 AlaGlnArgAlaAspMetLeuArgGlyLeuSer----- 1714
DB 22981 CGTAAAGCTCTTCCCATCCCGCGCGTGGTGGCTGCCCGTGAACAGCACCGCAA 22922
QY 1715 -----ProArgGluSerSerLeuAlaLeuAsnTyrgluAlaGlyProArgGlyIleIle 1732
DB 22921 GCTTTCGCGCTCGCTTCGCGCTCCGCTCACCGCGCGCATCCCGCGCGCGCGAGCA 22862
QY 1733 AspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAla 1752
DB 22861 GCGCGGAGAGACCTCCACAGTCTCCGCGCGCTCGCGCGAGCAGCAGC-----ACGCCCGCG 22805
QY 1753 ThrAlaMetAspArgLeuAlaTyrgluProThrAlaProGlnProPheSer----- 1769
DB 22804 ACTCGAAGTGGTGGTGCAGCGCGCGCTCTCTACACGCTCCGACACACCCACCTCCC 22745
QY 1770 ---SerArgHisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPro 1788
DB 22744 CGTGTCTTCCAGCCACTTCCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22700
QY 1789 ThrThrThrSerSerGluArgGluArgAspArgGluArgAspArgAspArg 1808
DB 22700 ----- 22700
QY 1809 GluArgGlyLysSerIleLeuThrSerThrThrValGluHisAlaProIleThrArg 1828
DB 22699 CCGAGGCTCGTCTCGCGCGCGAGCAACAGCGGTATCGCAACACACCTCCG----- 22649
QY 1829 ProGlyThrGlnSerSerGlySerSerGlySerGlyGlyGlyGlySerSer 1848
DB 22648 CTTCC-ACCGCTCGCGCGCGCGCTCCCGCGTTCGCGAGCGCGCTCCACCGCGCGCTCT 22590
QY 1849 SerArg-----ProAlaSerHisHisAlaHisGln----- 1859
DB 22589 TCAGAGTATGATGGGGTTCGCTCGCTGATCCGNAACAGCACCGCGCGCGCGCGCG 22530
QY 1860 HisSerProIle-----SerProAr 1866
DB 22529 GCCCGCGCTTGGCGCGCGCGCGCGCTCTTTCAGCAACAGCACCGCGCTCCCTCC 22470
QY 1866 gThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyI 1886
DB 22469 CACCCAAATGTGCGCGCTCGCTCTCGCTGCGAGCGGTTCGCGAGCACCTCGTGTGTC 22410
QY 1886 elleThrAla-----ValGluProSerLysProThrValLeuArg-- 1899
DB 22409 ATCAGACAGCACCATCTTTCAGCACACCGGCTACCGCGCGCGCGCGCGCGCGCG 22350
QY 1900 -SerThrSerThrSer-----SerProValArgProAlaAlaThr----- 1912
```

| | | | | | | | |
|----|-------|---|-------|----|-------|---|-------|
| Db | 26276 | TGGTCTGGACGCACTCCCGCGTCTCTCAGCGCGCTCCACGAGTCTCCACGAGCGCGG | 26217 | Db | 25220 | CCATAAAGCACGAACGATCCAGTCCGCGTCTCTCCAGCAGCAGTCTCTCGAGTGCCATGCT | 25161 |
| Qy | 709 | YValSerGly- ::: ::: ::: | 716 | Qy | 994 | -----ProAlaProAlaProProPro ::: ::: ::: | 1001 |
| Db | 26216 | TGCTCGGATCCATCGGCTCGGCTCGCGGGGCTTATCCGAAGAACCCTCGGCTCAAAC | 26157 | Db | 25160 | CCGTTGACCTTCGCCCGGAGCTCTCGCGCAGCATCTCGGGCGCGAGCTCTCGACCGGC | 25101 |
| Qy | 716 | uMetValGluAlaGluAlaLeuHisAlaSerGlyAsnGluValPro- | 732 | Qy | 1002 | -----ProGlnAsnLeuGlnProGluSerAspAla ::: ::: ::: | 1013 |
| Db | 26156 | GATGCCACGTCGTCACGAGCGAGCATGCTTACGTACGTCTTGCCCTTGCGCTCGGG | 26097 | Db | 25100 | ACGCGGGGAGCTATCCCGCCAGATGCCACCGCTCAGCGCGCTTCGTCTGTCTCA | 25041 |
| Qy | 733 | -----ArgGlyGluCysSerGly | 738 | Qy | 1013 | nGlnProGlySerSerProArg ::: ::: ::: | 1022 |
| Db | 26096 | TTAGGATCGTACATCGCTCCACGTCGAACGGTCGCTCGGATCGGCTTAGGATATCC | 26037 | Db | 25040 | AGTCTCTCACCAGCGCTCGACCTCGCTCGCTCGGCACATCGCAGCGCGCAAGCGTC | 24981 |
| Qy | 739 | -----ProAlaThrValAsnSerSerAspThr | 751 | Qy | 1022 | sSerArgSerProAlaProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe | 1042 |
| Db | 26036 | CGCCCTCCACGACACTTGCAGAGCTCTCCAGATCGTTCTCGCTCCGCCCGCATCGC | 25977 | Db | 24980 | ACGCGCGCGCCCGCCCTCAAGCTCTCCCGAAGCTCGCGCGCGCGCTCGCGGGCGG | 24921 |
| Qy | 752 | -----ProSerProHisThrGluAlaAlaLysAspThr | 767 | Qy | 1042 | uProGlyAspProCysTrpThrSerGlyLeuProPheProValProArgGluVal | 1062 |
| Db | 25976 | AGCCCCACGCGCATGCTATCGGCGCGCTCCGCTCCGCTCCAGCGCTCAGCGCGTGC | 25917 | Db | 24920 | CCCGGGGTGACGCCA-----GCACCAAGTGTCTCTCCCGCTCCGCCAAC-CAGCG | 24868 |
| Qy | 767 | oLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProAr | 787 | Qy | 1062 | lLeLysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaProProGly-- | 1081 |
| Db | 25916 | TCGTCGCGCGCGTCTCGCGCGCATCAGCTGCTCCAGCAGCGCGGTACCTCG | 25857 | Db | 24867 | AGCAGATGGCTCCCGAGCCCTCCACTCCACCCGTGATGAGCACCGTGCACGCGCTT | 24808 |
| Qy | 787 | gArgThrSer-----ArgAlaProIleGluProThrProAlaSerGluAlaTh | 803 | Qy | 1082 | -HisProLeuProLeuGly-LeuHisAspThrAlaArg | 1093 |
| Db | 25856 | CCTTCGCTCGGATTAATCGAAGATCAACGTCCTTGTGACCGACACCCCGCTCGCTGC | 25797 | Db | 24807 | CCACCCGCTCCCGCTCTCTCCACTCACCCGACGATCGCGCGCACAGACGCCACC | 24748 |
| Qy | 803 | xGlyAlaProProProProAlaProProSerProSer-----AlaProProValVa | 822 | Qy | 1094 | -----ProValLeuProArgProPro | 1100 |
| Db | 25796 | AGCGCTTCGACACTCACCGCCATCAACGAGTCCAGCGCGAGGTCCAGAACCTTCGC | 25737 | Db | 24747 | CGCTCGCAGCGCCACCTGGTCTCTCGTCTCTCGGAAACGAGCGTCTCTCAGCACCTGTG | 24688 |
| Qy | 822 | lProLysGluGlyGluGluGluThrAlaAlaAlaProProValGluGluGlyGluG | 842 | Qy | 1101 | --ThrIleSerAsnProProLeuIleSerSerAlaLysHisProSerVal----- | 1117 |
| Db | 25736 | TCCGGTCCAGCGCGCTCGATCTTCAGCGCGCAGACCGCGCGCTCGCTCGCCACG | 25677 | Db | 24687 | CACCGCTTCCCGCTCCACCTCCCGCGCGAGGTCTACAGTCCACCCAGCGCTCGGGTG | 24628 |
| Qy | 842 | uGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluProVa | 862 | Qy | 1118 | -----LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSer---- | 1130 |
| Db | 25676 | AGCTCGC---GCAGCCGCTCGCTCCGCTCCACAGCCGCGAGCGAGCATCTCCCG | 25620 | Db | 24627 | CTCAGGCCACACCGCGCCCAACCCACAGTCAGCGCTCGCAGGTGGACACCCC | 24568 |
| Qy | 862 | lLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGlu | 882 | Qy | 1131 | -----ValGlnLeuHisValProTyrSerGluHis----- | 1140 |
| Db | 25619 | A---CGGCTGTCCC-----CGGTCGCGCGCGGCTGTGCGGT | 25581 | Db | 24567 | TTGCTTTCCTCGCTGTCTACCGCTCTCGTCTCAACACACGAGCGCGCATCGAGCCC | 24508 |
| Qy | 882 | aAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGlyGlySerGly-- | 901 | Qy | 1141 | -----AlaLysAlaProVal---GlyProValThrMetGlyLeuProLeuProM | 1156 |
| Db | 25580 | GTCCGGCCCTCCAGCGCATGCGCGCTTGTCTCGATCCGTCACAGCGCGGCTCGCGCG | 25521 | Db | 24507 | AGCGTACCGAGCGCTCGCGCCAGCGCCAGCATCTGTACTCCGCGAGGCCCTCGCC | 24448 |
| Qy | 902 | -----ArgAlaThrAlaLysSerSerGlyAlaPro----- | 912 | Qy | 1156 | etAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgG | 1176 |
| Db | 25520 | GCCGCGCAGAACAGCGCGCGAAGCGGACAGTCCACGTCGCGATCGCCACCGAGGTG | 25461 | Db | 24447 | TTCTTCAGCGCACCTTGTCTCC-----CAGCGCGCTCAGCGCCACGAC | 24403 |
| Qy | 913 | -----GlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl | 929 | Qy | 1176 | lyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuA | 1196 |
| Db | 25460 | CGCGCAGCCCACTCGAGCCCGCAAGCGCTTGTCCGCGCATCGGCACCGACCGCG | 25401 | Db | 24402 | GCCACGAGCTCG-CCCTCCAGC-----CCTCTCAACCTCGCGCGCAGTGCCTC | 24353 |
| Qy | 929 | agluGlyGly-----AspLysAsnArgLeu | 937 | Qy | 1196 | rgGlyThrAlaLeuGlySerVal-----ProGlyGlySerIleT | 1209 |
| Db | 25400 | CGGCTCCGGAGCTGGGCTCGGCTCGTCTCACCATCCGCTCCGCGCTCCGAGCCAGCGGCC | 25341 | Db | 24352 | GCTCTTCGCTCGCGCTCGATGATACCTTCGACCCGAGCTCCGAGCGCATCTCTCA | 24293 |
| Qy | 938 | -----LeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe | 956 | Qy | 1209 | hrLysGlyIle---ProSerThrArgValProSerAspSerAlaIleThrTyrArgGlyS | 1228 |
| Db | 25340 | CAGTGACACCGTCCGCTGTGCTCGCGCAGCGCGGTAGCGCGGCGGCTCGAGC | 25281 | Db | 24292 | CCGCGCGGTTCAGCCCTCCGCGCGGAGATGACACGAGCAGC-----CATCTGC | 24242 |
| Qy | 956 | rProGlnLysProLeuAspLeuLysGlnAlaAlaAlaIleProIle | 976 | Qy | 1228 | erIleThrHisGlyThrProAlaAspValLeuThrLysGlyThrIleThrArgIleIleG | 1248 |
| Db | 25280 | CCTGGTTCGCGCGCGCTACCGCGCTCGCTCCAGAGCCCGACAGCCAGCGATGTG | 25221 | Db | 24241 | CCCTCAGCTCGGCTTCCCTCCCTG-----GCAGCGCTGGCTTCCCTTTGCC | 24197 |
| Qy | 976 | eGlnValThrLysValHisGluProProArgGluAspAlaIleProThrLys----- | 993 | Qy | 1248 | lyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLysGlyHisV | 1268 |
| | | ::: | | Db | 24196 | ACGCTCTCTGTATCAACAGCCAGACACCGTGGCTTCTCGGTCTCTCGCGGCCACG | 24137 |

```
QY 117 LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThr 136
Db 28089 CACCTCCAGCA-CGGCGCTTCGTCGCCGCTCAGCAGCGCTGCCTCG--GACCGTTGA 28034
QY 137 LysAspArgSerLeuThrGly-----LysLeuGluProValSerProProSerPro--- 153
Db 28033 TGCCCGCATCCCAAGCGCGCTCCAGCCCTATCGCGCCATCGCAGCGCCCTGCACCTCCG 27974
QY 154 --ProHisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluLeu 172
Db 27973 GCTCCGAGCTTCACCGACATCATCGCTCCGCCGCTCGCCTCGCCTCT--- 27926
QY 173 IleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLys 192
Db 27925 --GCATCAGCCGACCGC--GGCGCACACTAGCTTCGCTCGCGCAAGGCTCAGCA 27872
QY 193 LeuLysLysLysGlnGln-----LeuGluGluGluAlaLysProProGlu-Pr 210
Db 27871 CGCCCGCACGTCGCGAGCGCTCAGCTCTCTATCGAGTCCCGCAGCGCGCGCGCC 27812
QY 210 oGluLysProValSerProProIleGluSerLysHisArgSerLeuValGlnIleI 230
Db 27811 TCACTCCCAAGCGCTCCACTGCGGACAGCGCCTTCAG----- 27769
QY 230 eTyAspGluAsnArgLysLysAlaGluAlaHisArgIleLeuGluGlyLeuGlyPr 250
Db 27768 -----CGCAAA----- 27763
QY 250 oGlnValGluLeuProLeuTyraGlnProSerAspThrArgGlnTyrHisGluAsnI 270
Db 27762 -----CAACCC-----GGGCT 27752
QY 270 eLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgAsnHisAl 290
Db 27751 GCGTGACTCCGCTCCAGCGCCCTCCGCTCGCTCGCTCGCTCGCTTCGCGAACAACCA 27692
QY 290 aArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuG 310
Db 27691 CCTCTCTCAA-----CCACGCTCGA 27671
QY 310 uLysLysValGluArgIleGluAsnAsnProArgArgAlaLysGluSer----- 327
Db 27670 GATCGCGCTCCAGCGCTCCACACCTCGTCGAACGCGCGCACGAGGCTACACTT 27611
QY 328 ---LysValArgGluTyrGluLysGlnPheProGluIleArgLysGlnArgGluLe 346
Db 27610 CGTAAAGTCTCTCCCATCCGAGCGCTGGCTCCCTCGCTCCCGGTAACAGCACCGCAA 27551
QY 346 uGlnGluArgMetGlnSerArgValGlyGln-----ArgGlySerGlyLeu--- 361
Db 27550 GCTTCCCGCGCGCTCCGTCGCGCGCACACCTCGCGCGCTCGACCTCCGACA 27491
QY 362 -----SerMetSerAlaAlaArgSerGluHis----- 370
Db 27490 GCGCGCTCAGCGCTCCACCGCTCCGACACGCTCGCGCTCGACCGATCGCGCGCAGG 27431
QY 371 -GluValSerGluIleLeuAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetAr 390
Db 27430 CGAAGTCCGCTCGGTG-----CAGCGCGCGCG 27404
QY 390 gGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheI 410
Db 27403 TTCTCACCAGTCCGCGCACCGCACCTC----- 27376
QY 410 eAsnMetAsnGlyLeuMetAlaAspProMetLys----- 421
Db 27375 -----CGCGTGTCTCCGAGCCACTTCGCGCCAGCGCTCGCGCTCGCGCCCGCCACCG 27326
QY 422 ---ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
Db 27325 AGGCTCTGTCTCGCCCGACAG-----CAACAGCGGGATCGGACACCTT 27281
QY 440 eArgGluLysPheMetGlnHis----- 447
Db 27280 CGCGCCCAACGCTCGCGCAGCGCGCTCGCGCACCGCTCAAGCGCGCGCTCTT 27221
QY 447 ----- 447
Db 27220 CGAGGATGATGCGCGCTTCGTCCTCGCTGATGCAAAACGACGACGCGCGCAAGCGGG 27161
QY 448 -----ProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysTh 462
Db 27160 GGAGGCGCTCGCGCGCGCGCGCGCTCAGCGCTCAGCGCTCGACCTCGACCGCGGAC 27101
QY 462 rValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLe 482
Db 27100 TCTCCCGAGGATCTG-----GGGCT 27080
QY 482 uValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln 501
Db 27079 TGCGCGGAGCTGTCAACGTCGCGCGGCGGCTCATGCCGAAACGCGCGCAGGATCT 27020
QY 502 -----GlnGlnG 504
Db 27019 TGCACACTCGGCGATCGCGCGCGCGCTCCAGGTGACCGATGTTGCTCTTTCACCGCTC 26960
QY 504 nGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlnGluLysAspGluLysG 524
Db 26959 CAAAGCGCAGCGCGGAGTCCCTCGCTTGCCTGCTACACCGCGCGCGGCTTGCA 26900
QY 524 uLysGluLysGluAlaGluLysGluGlu----- 534
Db 26899 CTTGATCGATGCGCGCGCGCGCTCGCTCGCTGATGATGACCGCGCTCGCCCGCACCGC 26840
QY 535 ---LysProGluValGluAsnAspLysGluAsp----- 544
Db 26839 ACCTCTCAGCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCTTCTGCTGGCGCGCTCGCT 26780
QY 544 ----- 544
Db 26779 TCGCGCGCTGATCCGCTGCTCGCGCGCTGCTGATGATGACCGCGCTCGCCCGCACCGC 26720
QY 545 -----LeuLeuLysGluLysThrAspThrAspThrSerGlyGluAs 557
Db 26719 CCAGCACCGCATCCCTCGCGCTGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26666
QY 557 pAsnAspGluLysGluAlaValAlaSerLysGlyArg-----LysThrAlaAsnSe 574
Db 26665 CCGCGAGCGCTCGCGCGCGCGCTAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 26606
QY 574 rGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluG 594
Db 26605 CGTCGCGCGAGACCGCA--CGCTGCGCGACAGCAGCAGCAAAACCGCGCGCTTCGCGAG 26548
QY 594 uAlaIleThrProGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerAr 614
Db 26547 CACTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26489
QY 614 gTptThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTr 634
Db 26488 CCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26430
QY 634 pSerAlaIleAlaArgMetValGlySer-----LysThrValSerGlnCysLysAsnPh 652
Db 26429 TGCAGCGCGGAGATGATGACACACCGCGCGCTGCGAAGCTCAGCGCGCTCGCGCTCAGC 26370
QY 652 eTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLy 672
Db 26369 GCATCGCATCTTCGTCGCGCTCTTCGCGCGATAGC----- 26333
QY 672 sMetGluLysGluAsnAlaArgArgLysLysLys-----LysAlaProAlaAl 689
Db 26332 ----TCGCATCTCTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26277
QY 689 aAlaSerGluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGI 709
```

Qy 2242 TyrProLeuLeuTyrArgAspGlyGluGlnThrCluProSerArgMetGlySerLysSer 2261
 Db 1034 ATGTTGCTTGTGCACAGAGGGAGTGGAGCCCTGCTGAGCAAGAGTGTATCTCGATCA 1093
 Qy 2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
 Db 1094 CCAGGAAGTAACTACTTGCCTTCATCTTCCACCAAGCTT---GAAAGCACATCACCC 1150
 Qy 2282 MetValLysSerLysGlnGluLeuLeuAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
 Db 1151 ATGCTTAAATCAAGAAACAGAGAAATTTTCGTAAGTTGAATCTTCTGCTGGAGGTGAC 1210
 Qy 2302 ProGluTyrAsnLysSerGlnProGlyThrGluLeuPheAsnMetProAlaIleThrGly 2321
 Db 1211 TCTGATATGGCAGCTGCTCAGCCAGGAACAGAGATCTTCAATCTGCCAGCAGTTACCACA 1270
 Qy 2322 ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
 Db 1271 TCAGGTGAGTGAAGCTCAAGAGCCATCTTTTCTGTATCCGCCAGT---AACCTTGGT 1327
 Qy 2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnThrGluGlu----- 2359
 Db 1328 CTAGAAGACATCATCAGAAAGCTCTCATGGAAGTTTGTATGATAAAAGTTGAAGATCAT 1387
 Qy 2360 -----SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
 Db 1388 GGTGTGTCATGTCCTCCATCTCTGTG-----GGCATTTATGCTGTGTAGTGGCAGCAC 1438
 Qy 2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
 Db 1439 TCAGTG-----GTGACGAGCAGCGAGGACGAGAGATGAAGGGGAGCCA 1483
 Qy 2396 SerProGlyGlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
 Db 1484 TCACCTCATGAGGAGTATGCAAAACCAAGCTGATCAACAAATCAAAACAGCAGAACTCT 1543
 Qy 2415 LysSerProAlaProGlyLeuAla-----SerGlyAspArgProProSerValSerSer 2432
 Db 1544 AATCTCTCTATCTCTGGGCAAGCTATTATGAACTGAAGGCTTCTTCTGCTCTCTCT 1603
 Qy 2433 ValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAsp 2452
 Db 1604 GTGCATTCAGAAAGTGATTACACAGGCGACACCA-----GGATGGCGATCGGAAGAT 1657
 Qy 2453 ArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGln 2472
 Db 1658 CGGCCCTCTTCAACAGGTTCTACTCAGTTCCCTTACAAACCTCTGACCATAGGATG--- 1714
 Qy 2473 AlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu--- 2491
 Db 1715 -----CTCAGCAGTACACCTACACAGATCGCATGGCCCATCTGCCATCACC 1765
 Qy 2492 ---AlaGlyProHis-----HisAlaTrpAspGluProLysProLeuLeuCys 2507
 Db 1766 CAAGCAGCTCCACATCAACAGAACCGCATCTGGAGAGAGGAGGCTGCCCCGCTCTCTCA 1825
 Qy 2508 SerGlnTyrGluThrLeuSerAspSerGlu 2517
 Db 1826 GCGAGTATGAGACACTGTCTGATATGAC 1855

RESULT 23

ADC26995/c
 ID ADC26995 standard; DNA; 67251 BP.
 XX AC ADC26995;
 XX AC ADC26995;
 DT 18-DEC-2003 (first entry)
 XX Sorangium cellulosum tmbA gene cluster.
 XX recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;
 KW polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;
 KW ds.

XX Polyangium cellulosum.
 OS US2003054547-A1.
 XX 20-MAR-2003.
 XX 28-AUG-2001; 2001US-00942025.
 XX 22-JAN-1998; 98US-00010809.
 PR 31-AUG-1998; 98US-00144085.
 PR 15-FEB-2001; 2001US-0271245P.
 XX (JULI/) JULIEN B.
 XX Julien B;
 PI WPI; 2003-567100/53.
 DR XX
 XX New recombinant DNA vector for preparing hybrid polyketide synthase
 PT enzymes comprises all or a portion of one or more genes in the tmbA gene
 PT cluster of Sorangium cellulosum.
 XX
 XX Disclosure; SEQ ID NO 1; 73pp; English.
 PS
 XX The invention describes a recombinant DNA vector that comprises all or a
 CC portion of one or more genes in the tmbA gene cluster. Also disclosed is
 CC the host cell that comprises the tmbA genes of Sorangium cellulosum or
 CC fragments of those genes. The DNA vector is useful in preparing hybrid
 CC polyketide synthase (PKS) enzymes and the polyketides produced by such
 CC hybrid enzymes. The gene products of the tmbA cluster can be used to
 CC synthesize the polyketide tombamycin. This sequence represents the
 CC Sorangium cellulosum tmbA gene cluster.
 XX
 SQ Sequence 67251 BP; 9090 A; 19855 C; 27880 G; 10420 T; 0 U; 6 Other;
 Alignment Scores:
 Pred. No.: 1.27e-12 Length: 67251
 Score: 708.50 Matches: 665
 Percent Similarity: 30.74% Conservative: 271
 Best Local Similarity: 21.84% Mismatches: 1189
 Query Match: 5.36% Indels: 933
 DB: 9 Gaps: 132
 US-09-522-753-5 (1-2517) x ADC26995 (1-67251)
 Qy 2 SerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProPro 21
 Db 28456 GCAGGCTCGCACAA-----ACGTGCTCGGCTCCCTCCG 28421
 Qy 22 HisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGlu 41
 Db 28420 GCAGGACCCCGCCCGCGCGCACAGCAGC---CATCCGGAC----- 28379
 Qy 42 TyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnPro 61
 Db 28378 -----CGCACTCGACGTATGTCTACCCCGCGCGCGCA-ACCGTGGCATCCCC 28329
 Qy 62 GlnArgArgArgPro-----SerLeuLeuSerGluPheGlnProGlyAsnGluArg 78
 Db 28328 TCCAGGAAGCGCACCGCTCGCGCAGCTGCTCACCAGTACTCGGCGCA-CATCAGGA 28270
 Qy 79 -----SerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeu----- 92
 Db 28269 TTCTTCGCCACCGAGCTCGCGCTCACCGCGCTCACCGGCTACCTGCGGCTTCCATA 28210
 Qy 93 -----ProGluLeuGlyLysSer-----GluMetGluPheIle 103
 Db 28209 CGCGACTTCCCGGCCACCTCCCGTACTTTCAGCATCCCGTCCATGTGCGCGCTGTG 28150
 Qy 104 GluSerLysArg-----ProArgLeuGluLeuLeuLeuProAspPro 116
 Db 28149 GAATCGTGCACACGCGCGCAGCGCGCGCTGCGCGCTCGAACCCTGCC 28090

Db 4262 CCTCTGCGGTCGGGA-GCTTCTGCGGCGCGCAGCAGCGGATGGCGGCCGCCCTTCAG 4204
 QY rpGluGluSerProProLeuSerAlaAenAlaAheAenProLeuAenAlaSerAlaSerL 2377
 Db 4203 CCCGCGGACCCACGTCACCGCGGCTCTCGGCGCTCCAGCAGTAGGCTGCC--- 4148
 QY 2377 euProAlaAlaMetProIleThAlaAlaAspGlyArgSerAspHisThLeuThSerP 2397
 Db 4147 -----GCTGGAATCGGTACGTCGCGCAGCTCCACGGCGCGCGCCATGGCCGACAGCAC 4093
 QY 2397 roGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerP 2417
 Db 4092 CTGGG-----CCGAGC 4081
 QY 2417 roAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluG 2437
 Db 4080 CACTCGTGCCTTGCACGTGCACCGTCCGACCGCC-GTTGCAAGCGCGCTCT---T 4025
 QY 2437 lyAspCysAenArgArgThProLeuThrAsnArgValTrpGluAspArgProSerSerA 2457
 Db 4024 CTTCTGCTCTCGCGCAGGCTCGCCACAAACGTCGCC-----TCGGCTCCCT 3977
 QY 2457 laGly-SerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet 2476
 Db 3976 CCGCAGGACCCCGCCCCCAGCGCGCAGCAGCGCGTCCGCGACCGCCTCGACGTATG 3917
 QY 2477 AlaSerProProProGlyLeuProAlaGly-----SerGlyProLeuAla 2492
 Db 3916 TGGTACCCCGCGCGCGAAGCGTGGCATCCCGTCGAGAACCTCACCGCTCTCGCG 3857
 RESULT 22
 AAT31931
 ID AAT31931 standard; DNA; 1922 BP.
 XX
 AC AAT31931;
 DT 27-SEP-1996 (first entry)
 DE Retinoid X receptor interacting protein RIP13 DNA.
 XX Retinoid X receptor interacting protein; RXR; RIP; RIP13; ss.
 KW Retinoid X receptor interacting protein; RXR; RIP; RIP13; ss.
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 2..1861
 FT /*tag= a
 XX
 PN W09621677-A1.
 XX
 PD 18-JUL-1996.
 XX
 PF 08-DEC-1995; 95WO-US016311.
 XX
 PR 13-JAN-1995; 95US-00372652.
 XX
 PA (GEO) GEN HOSPITAL CORP.
 XX
 PI Moore D, Seol W, Choi H;
 XX
 DR WPI; 1996-342241/34.
 DR P-PSDB; AAR99738.
 XX
 PT Retinoid X receptor (RXR) interacting protein (RIP) - useful to modulate
 PT or mediate RXR function, anti-RIP antibodies can be used to determine RIP
 PT subcellular distribution patterns.
 XX
 PS Claim 8; Page 56-57; 90pp; English.
 XX
 CC A cDNA clone (AAT31931) codes for retinoid X receptor interacting protein
 CC RIP13 (AAR99738), which is a candidate transcriptional co-activator. It
 CC was isolated from a mouse liver cDNA library. The cDNA can be used for
 CC the recombinant prodn. of RIP13 in transfected host cells

XX
 SQ Sequence 1922 BP; 545 A; 508 C; 451 G; 418 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4, 22e-14 Length: 1922
 Score: 718.00 Matches: 212
 Percent Similarity: 48.81% Conservative: 76
 Best Local Similarity: 35.93% Mismatches: 214
 Query Match: 5.43% Indels: 88
 DB: 2 Gaps: 25
 US-09-522-753-5 (1-2517) x AAT31931 (1-1922)
 QY 1975 ProSerLysGlySerGluProArgProLeuVal----- 1985
 Db 209 CCAAGTGGCAAGGCCCGCAGCTCATGCTCAGTAGTGTATTCTGAGCTGGAAAGATAA 268
 QY 1986 -----ProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAen----- 2001
 Db 269 GGGCTCTCTCCAAATCCAGATATGAGGAAGAGCTTAAGGACCCGAGGGAAGACTACATT 328
 QY 2002 LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPro 2021
 Db 329 ACTGCAGCTAACTTCATAGCAGTGCACATCCCGCAAATTCCTTCGGACAGGATGCG 398
 QY 2022 HisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeu 2041
 Db 389 AGGGAACGTGGCTCTCAAAGTTCCAGACTCTTCT-----AGTAGCTTG 430
 QY 2042 GlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSer 2061
 Db 431 TCTTCTCACAGGTATGAAACGGCTAGTAGTGCATTCAGGTGATAAGTCCCGCAGCTCA 490
 QY 2062 ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHis 2081
 Db 491 CTGCACACCCAGGAAAGCCACAGGCTATCAGCCAGCATGTTTAAGGCAATCAA 550
 QY 2082 LeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAla 2101
 Db 551 GCAGAAATGAGTCCACTCGACAGTATGAAGGTCCA----- 586
 QY 2102 HisLeuProHisLeuArg-----ProLeuProGluSerGlnProSerSer 2116
 Db 587 ---CTGCATCATTCGTCGCCAGCAGGAATCACCATCTCCACAGCAACAGCCACCTG 643
 QY 2117 SerProLeuGlnThr-----AlaProGlyValLysGlyHisGlnArgValThr 2134
 Db 644 CCCCATCTCCAGTCAGAGGAATGGACAGTCCCGAGGCCCATCGACTGATCACA 703
 QY 2135 LeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGln 2154
 Db 704 CTTGCTGACACATCTGTCAAATATATCACAAAGATTTTGTAGAAAT-----CAAGTT 757
 QY 2155 LeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAla-----SerCysPro 2172
 Db 758 CCTCGCAGCTCTTACTTCTACATTCCAAACCTTCCATCTGCTTGTTCATCCACCT 817
 QY 2173 ValLeuAspLeuArgProProSerAspLeuTyrLeuProProAsp----- 2189
 Db 818 GTA-----AGAACTAAACCTCAAGCCGTACGCCCAAGTACAGTCTCAGACT 868
 QY 2190 -----HisGlyAlaProAla---ArgGlySerProHis-----SerGlu 2201
 Db 869 GTCTTCATCCCGACAGCAGTCTCAGAGTCTCTCCAGAAATCTTGTGGATAAATCCCG 928
 QY 2202 GlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGly 2221
 Db 929 GGAAGCAGCGCTGGAAAAATCTCCAGAGAGAGTGCATATC-----CCATCAGAGCCC 979
 QY 2222 IleGluProValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
 Db 980 TATGAGCCCATCTCCCAACCCCAAGGCCCTGCTGTG-----CATGAGAGCAGGACAGC 1033

[illegible]

| | | | |
|----|------|--|------|
| Db | 9166 | CGGACATCAGCGCTTCTTCCGACCGAGCTCGCGGCTCAGCGCGTCAACCGCCGAGCT | 9107 |
| Qy | 898 | GlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSer | 917 |
| Db | 9106 | CGCGCGCGGTACGTACGTCTCCCGCGCCACCTTCCGAACTCTCTCCAGCATCCCGTCCA | 9047 |
| Qy | 918 | SerAlaThrCysSer-----AlaAspGluValAspGluAlaGluGlyGlyAspLysAsn | 935 |
| Db | 9046 | TGTGCGCGCTGTGGAAACGGGTGGACACACGACGACGCGCGCTGCGCGCGCTGCGCCT | 8987 |
| Qy | 936 | ArgLeuLeuSerPro----- | 940 |
| Db | 8986 | CGAGCGTCCGCGCAACCGCGAGCACCGCGCTTCGTCCCGCTCAGCACCGTCTGCATCG | 8927 |
| Qy | 941 | -----ArgProSerLeuLeuThrProThrGlyAsp--- | 950 |
| Db | 8926 | CGCGGTTCAGCGCGGCGATGCTCAGTCGCGCTTCGCGCGCGAGCTCCGACAGCGCGCT | 8867 |
| Qy | 951 | -----ProArgAlaAsnAla | 955 |
| Db | 8866 | GCACTCCGCGCTCGAGGCTCCACGACACCATCGTCCCGCGCTCGACCCGTGCA | 8807 |
| Qy | 956 | SerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaLalleProPro | 975 |
| Db | 8806 | TCAGCGACCGCGGGCGCACACTAGCTTCGCTCGCTCCGCAAGGCTCAGCACACCGCCCA | 8747 |
| Qy | 976 | -----IleGlnValThrLysValHis | 982 |
| Db | 8746 | CGTCCGAGCGCTCAGCTCTCTATCGAGTGCCTCAGCGAGCGAGCGGTCTCAGCCCCC | 8687 |
| Qy | 983 | GluProProArgGluAspAlaAlaProThrLys----- | 993 |
| Db | 8686 | ACGACTCCCACTGACGCTACAGCGCCACTTCCAGCGCAACAGCCCGGCTGCGTGTA | 8627 |
| Qy | 994 | -----ProAlaProProAlaProProProPro | 1002 |
| Db | 8626 | CCGTCCGCTCCAGCAACGCTCTCTCTCGCTGCGCGCGAGCGCGGCAACACACCTCCGCA | 8567 |
| Qy | 1003 | GlnAsnLeuGlnProGluSerAspAlaProGlnProGlySerSerProArgGly--- | 1021 |
| Db | 8566 | GACCACCGGAGGTGCGGCTCAGCGCTCAGCACCTCGTGAAGCGCGCACGNAACA | 8507 |
| Qy | 1022 | -----LysSerArgSerPro----- | 1026 |
| Db | 8506 | CGGGGTACACTTCGTAAAGCTCTTCCCATCCGAGCGCTGGCTGCCCTGCCCGCTGA | 8447 |
| Qy | 1027 | -----AlaProAlaAspLysGluAlaPheAlaGluAla----- | 1039 |
| Db | 8446 | ACAGCACCGCAAGCTTCCCCCGCGCACTCCGTCGCGCGACACACCCCGCGCTGCTC | 8387 |
| Qy | 1040 | -----GlnLysLeuProGlyAspPro | 1046 |
| Db | 8386 | GACCTCCGACGAGCGCTCAGCCCTCCACAGCTTCCGAGCGCTCGCGCAAGCACCG | 8327 |
| Qy | 1047 | -----ProCysThrThrSerGly-----LeuProPhePro | 1056 |
| Db | 8326 | ACGCGCGGTGGTAAAGTGGTTCGCTGCGAGCGCTGCGTCTCTCACCGTCCGACCCACC | 8267 |
| Qy | 1057 | ValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSer | 1076 |
| Db | 8266 | CCACTCCGCGTGTCTTCCAGCCACTTCGCGCGAGCGCGCGCT----- | 8222 |
| Qy | 1077 | TyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla----- | 1092 |
| Db | 8221 | -----GCGCGCGCACCGAGCCT-----GGTCTCGCGCGACAGCAACAGCGGCATCG | 8174 |
| Qy | 1093 | -----ArgProValLeuProArgProProThrIleSerAsnProProLeuIleSerSer | 1111 |
| Db | 8173 | CTGACCCCTCTGCGGCTTCGCGCGCACCGGCTCGCGCACCGCTCCA----- | 8126 |
| Qy | 1112 | AlaLysHisProSerValLeu-GluArgGlnIleGlyAlaIleSerGlnGlyMetSerVa | 1131 |

Db 12308 GCCTCCAGCGCGCAGGAGGC-----CGTCGATCGCGTCTTTCCCTCGCGCCAAAGAGA 12255
QY 88 rHisSerTyrLeuPro-----GluLeuGlyLysSerGluMetGluPheIleGluSe 105
Db 12254 CGCAGTAGTCTTCGGCGCTCGACGCTCTCCGGCAGCGCGCACGCCATCGATCAGATC 12195
QY 105 rLysArgProArgGluLeuLeuLeuProAspProLeuLeuArgProSerProLeuLeuAl 125
Db 12194 GCCACGCGCTCTTCCTCCCGCATGACGAGCGCGCGCTTCCTTCACT---GCAGC 12138
QY 125 aThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLe 145
Db 12137 CCGAGAACCGCTTCGCTTCAGCAGAGCTG---CGATCGCTCGCGGTGTCGGGTAGTCG 12081
QY 145 uGluProValSerPro-----ProSerProHisThrAspProGluLeuLeuLeuVa 163
Db 12080 AAGACCAGCTCGCGCGCAAGGAGCTTCGCGCGCGCGGCGGCGGTTCGCGAGCTCC 12021
QY 163 lProProArgLeuSerLys-GluGluLeuIleGlnAsnMet----- 176
Db 12020 ACCGCCATCAGCAGTTCGAGCGCGAGCGCTTCAGCACCTGCTCTGCCCGCCAGCGCTCG 11961
QY 177 -----AspArgValAspArgGluIleThrMetValGluGlnGlnIleS 191
Db 11960 CTCGCTGGAGCCCGAGCACCGCGCGGACCTCGCCCGCACCGAGCTCGATGAGCGCATTC 11901
QY 191 erLysLeuLysLysGlnGlnGlnLeuGluGluAlaLysProProGluProG 211
Db 11900 AGCGCTCCGCTCCGCGCAGCGGA-----GAGCGCGCGCGCAGC 11859
QY 211 lLysProValSerPro----- 216
Db 11858 GCCAGCGCTCGCTCGTTCGCTGCGTTCGAGCGCTTCGCAAGCTGGCGCGCAAC 11799
QY 217 -----ProProlleGluSerLysH 223
Db 11798 AGCGCAGAAACAGCGCGCAGCTCGCGCTGAGTTCACGCCCGCAGCTCGAGTCGCGG 11739
QY 223 isArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu----- 239
Db 11738 AG-----ATCCAGTGCATGGGACCGAGCTCGTTCGGAGCGCAAGAGCGCGCG 11688
QY 240 -----AlaAlaHisArgI 244
Db 11687 TCCAGCAGAGCGAGCGCTCTTCACCGCGCATCGGCAGCAACCGTTCGCGCTCATCGCG 11628
QY 244 leuGluGlyLeuGly-----ProGlnValGluLeuProLeuTyrAsnG 259
Db 11627 GACAGCTCGCGCTTCGCCAGGTGCGTGTCTATCCCCACGCCCGCTG-----T 11580
QY 259 lnProSerAspThrArgGlnTyrHis-----GluAsnI 270
Db 11579 GTCCAGAGGCCCGCAGCGCTCGTCCGCGGAGCGCGCGCGCGCGCGGTGGCGCGG 11520
QY 270 leLysIleAsnGlnAlaMetArgLysLysLeuLeuLeuTyr-----PheLysArgA 287
Db 11519 AGCGCGTCGAGGAACGTTTCGCGCGCGTGTCTGCGCGCGCGTTCGCGCGCGCAACGCTG 11460
QY 287 rGlnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGlnLeuMetG 307
Db 11459 CCGGCCACCGCAGAGACGACGACGCGCGGAGATCCAGCTCCCGCGTCCGCTCGTGC 11400
QY 307 luAlaLeuGluLysLysValGluArgIleGluAsnAsnPro-----ArgA 322
Db 11399 AG---GTGCGAGCGCCCATCCACCTTCGCGCGGACACCGCTGAAATGCGCTCGCGCTC 11343
QY 322 rGArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgL 342
Db 11342 TGCAGCACAGGACGCGCTGCTCAGCAGCGCGCGCAGTGCAGCACCGCGCTCAGCGGG 11283
QY 342 ysGlnArg----- 344
Db 11282 CGCGCGCGTTCGATGCGCGCGCAGCAGCGCGGACCTCTCCCGCTTCGACAGCTCGCAC 11223

QY 345 -----GluLeuGlnGluArgMetGlnSerArgValGlyG 356
Db 11222 GCAGCCACCGTCAACGCTCTCGCGCGCGAGCTCGCGAGCGATTCACAAGCTCGCG--- 11167
QY 356 luArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleI 376
Db 11166 -----GGCCCCGGGCGCTTCACGCGCGCGCGACGTCAACACAGGTCCTCAGCGCG 11112
QY 376 leAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArg-----GlnLeuAlaV 394
Db 11111 TCGCGCGG-----CACCAGGTGCGCGCGCGCGCTGCCCC 11076
QY 394 alilePro---ProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetA 413
Db 11075 AGTCCCTCCGTTCGCGCGGTCAACAGCAGCGTCCCGCAGGTCGACGTCCGCGGTCTCT 11016
QY 413 snGly---LeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetI 432
Db 11015 CGSGTGAGCTCTTCGCTCCGCTGACGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCC 10956
QY 432 rpSerGluGlnGlu-----LysGluThrPheArgGluLysPheMetGlnHisProLysA 450
Db 10955 CGCGCAGCGCGAGCTCCGCTCCGCGCGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10896
QY 450 snPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrT 470
Db 10895 TCCACGCGCTC---GGTCCCGCATCATCAAGCGCAGCGCGCGCTCGGGGTCTCGCTT 10839
QY 470 yrTyrLeuThrLysLys-----AsnGluAsnTyrLysSerLeuValA 484
Db 10838 CGCGCGTACGAATAAGCCCCACAGCGCGCGTTCGCCAGGTCTCTCGACGCGCTGTTTC 10779
QY 484 rGArgSerTyrArgArgGlyLysSerGlnGln-----GlnGlnGlnG 499
Db 10778 GCGCGCGCGCGCGCGCGCTCGCTCGCTATCCACCAAGCTCGACATCGTCAGCGCGCGG 10719
QY 499 lndGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 511
Db 10718 TCCGACAGCAGCGCTTCGACGACGACGACGCGCTCGCTCGCTCGCTCGCTCGCTCGCT 10659
QY 511 ----- 511
Db 10658 GCCAGTCCAAACGCGTTCGCGGTTCGCGCGGTGACGTTCGACACACCGCTCAGCGCGG 10599
QY 511 ----- 511
Db 10598 CTCGCCCATGCTCGAGGCGCGCAACCATGATCGATCGAGATCGCAATCGCTCGCGCCCC 10539
QY 512 -----MetProArgSerSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 526
Db 10538 AGSGCTTCGCGCGCGCTCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 10479
QY 527 ---LysGluAlaGluLysGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 543
Db 10478 TCCGGAGCGCTCGCCACCGCTTCACCGGTTCGAGTTCACCGCGGTACACCGCTCGGGCA 10419
QY 544 -----AspLeuLysGluLysThrAspThrSerGlyGluA 557
Db 10418 TTGGCAGCGTGGCCACTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 10359
QY 557 spAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyA 577
Db 10358 AC-----GCTGCCACACGCTCGCGCGCGCTCGCGCGCGCTCGCGCGCGCTCGCGCG 10314
QY 577 rGArgLysGlyArg-----IleThrArgSerM 586
Db 10313 GCAACCGCTTCGCTCATGTGCTTCAAGCTCAAGCATCACCGGAGCTCCCGCGGTCCC 10254
QY 586 etAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln-----GlnSerAlaGluL 604
Db 10253 GTCCGAGCAGCGCCAC-----GTCCGACACGAGAGGCGCAGCGCCATG 10206

```
Db 15199 TCAGGAGATCTTCGAATCGGAATCGGAACCGGACCTCGACGGATGTTTCGGACGGC 15140
Qy 2260 ---LysSerProGlyAenThrSerGlnProProAlaPhePheSerLysLeuThrGluSer 2278
Db 15139 CCGCGAGCTCCAAGAGCTTGGGAGGAGCGCTGAACGACGCGACGAGACTGACGGAGG 15080
Qy 2279 AsnSerAlaMetValLysSerLysGlnGluLeuAsnLysLeuAsnThrHisAsn 2298
Db 15079 CTGAGCATGAACGCGCAATTTCTCTCGACATCGAGTCCGCTGTACGACGAGATTC 15020
Qy 2299 ArgAsnGluProGluThrAsnLysSerGlnProGlyThrGluLeuPheAsnMetProAla 2318
Db 15019 CGCGCTCACGTTCCGAATTTGACGAGGAGAC-----TTCGGGATGTCGGTG 14969
Qy 2319 IleThrGly-----ThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db 14968 GCAGATGATGCTCGAAAGCAAAAGTGGCAGCAAGTTGCTTCGACCAAGAGCGAGC 14909
Qy 2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGly 2352
Db 14908 CGATTGCGGACTTCGACGGCCATG-----14885
Qy 2353 LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAlaPheAsnProLeuAsn 2372
Db 14884 -----ACGGAATCCATTCCGAGCTCCCTCAAGGGCTGGCGGTAGGAATC 14840
Qy 2373 AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaPheGlyArgSerAspHis 2392
Db 14839 GCGTCA-----CGCGCTGAAGCCCGAGCACCAGCGGACCTCGGCCCGCACAAAG 14789
Qy 2393 ThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArg 2412
Db 14788 TCGATGACGAGCTC-----AGGCGCTCCGCTCCGCGACGCGCGAGAGCGC 14741
Qy 2413 LysAlaLysSerProAlaProGlyLeuAlaSerGlyAsp-----ArgProProSerVal 2430
Db 14740 TCGCGGAGCTCGACGCTCTCTCTCGCGCAGAGCAGCGCTTGCAGCGCGGCGCGC 14681
Qy 2431 SerSerValHisSerGluGly-----AspCys 2439
Db 14680 ACCAGCGCGGAGAGCGCGCGAGCTCGCGCTGGAGCTCCAGCCCGAGCTGCAGCTGC 14621
Qy 2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
Db 14620 GCGAGATCCAGGTGCATCGGACCAAGCTCCTTCAGAGCTGAGAGCGCGGCTCCAGC 14561
Qy 2460 ThrProPheProTyrAsn-ProLeu-----IleMe 2469
Db 14560 AGAGCGAGGCGCTCTCGACCGATATCGGCAGAGCGCTCGCGCTGATCGCGGACAGT 14501
Qy 2469 tArgLeuGlnAlaGlyValMetAlaSerProPro-----ProGlyLeuProAla-- 2486
Db 14500 TCGGCGCTCGCCAGCTGGCTGTGTATCCCAAGCGCGCTTGCAGCCCAACCCCGCAGCC 14441
Qy 2487 -GlySerGlyProLeuAla 2492
Db 14440 AGGCTCGTCCGCGGAGCC 14422

RESULT 21
ID ADC26979/c
XX ADC26979;
XX AC ADC26979;
XX DT 18-DEC-2003 (first entry)
XX DE Sorangium cellulosum tmbA gene cluster tmbA DNA.
XX KW recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;
KW polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;
XX ds.
```

```
OS Polyangium cellulosum.
PH Key Location/Qualifiers
FT CDS 1..27705
FT /*tag= a
FT /product= "tmbA"
XX US2003054547-A1.
XX 20-MAR-2003.
XX 28-AUG-2001; 2001US-00942025.
XX 22-JAN-1998; 98US-00010809.
XX 31-AUG-1998; 98US-00144085.
XX 15-FEB-2001; 2001US-0271245P.
XX (JULI/) JULIEN B.
XX Julien B.
XX WPI; 2003-567100/53.
XX P-PSDB; ADC26980.
XX New recombinant DNA vector for preparing hybrid polyketide synthase
XX enzymes comprises all or a portion of one or more genes in the tmbA gene
XX cluster of Sorangium cellulosum.
XX Disclosure; SEQ ID NO 12; 73pp; English.
XX The invention describes a recombinant DNA vector that comprises all or a
XX portion of one or more genes in the tmbA gene cluster. Also disclosed is
XX the host cell that comprises the tmbA genes of Sorangium cellulosum or
XX fragments of those genes. The DNA vector is useful in preparing hybrid
XX polyketide synthase (PKS) enzymes and the polyketides produced by such
XX hybrid enzymes. The gene products of the tmbA cluster can be used to
XX synthesize the polyketide tombamycin. This sequence represents the
XX Sorangium cellulosum tmbA gene cluster tmbA open reading frame.
XX SQ Sequence 27705 BP; 3621 A; 8056 C; 11807 G; 4214 T; 0 U; 7 Other;

Alignment Scores: 2,22e-13 Length: 27705
Pred. No.: 723.50 Matches: 696
Score: 31.61% Conservative: 311
Percent Similarity: 21.85% Mismatches: 1209
Best Local Similarity: 5.47% Indels: 978
Query Match: 9 Gaps: 146
DB:

US-09-522-753-5 (1-2517) x ADC26979 (1-27705)
Qy 17 ProArgTyrProHisSerLeuSerTyrProVal-----Gln 29
Db 12608 CCGCGTACCTGCTCCAGCGCTTCAGCCCGCGGTATGAGAGCAGCAGCTAGTCGAGCCC 12549
Qy 30 IleAlaArgThrHisThrAspValGlyLeu-----LeuGluTyrGlnHisHisSer 46
Db 12548 ATGACCCGAGATACACTCCGCTGGCGCTCCGCTCAGCGCGCGCGCAGCGCGCA 12489
Qy 47 ArgAspTyrAlaSerHisLeu-SerProGly-----56
Db 12488 CGCTCCAATGCTCGCAGCGCGCTCTCCAGCAGCAGCGCGCTGCTGGGATCCATCGACTGC 12429
Qy 57 -----SerIleIleGlnProGlnArg-----63
Db 12428 GCCTCGCGCGCGATATCCCGAAGAGTCCGCTCGAAGAGTTCCTCCGCAAGAT 12369
Qy 64 -----ArgArgProSerLe 68
Db 12368 CCACCTCGCGCGGTAGCTCTTGGCCACCGCGCTCCGAGTCGGGGTCTAGACCGAAAGC 12309
Qy 68 uLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSe 88
```

```
QY 1655 euAlaProAsnProThrTyrPro-----HisLeuTyrProProTyrLeuIleArgGlyT 1673
D 17135 CCGCTCCAGCGAGCGGTGCTCTCTCCCGCATCTCCGCGCAGCACCGACAGACCA 17076
QY 1673 yrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerG 1693
D 17075 GCCCGTCCACGGCGCAGCGCCCTCATCGCAGCGCTGCA----- 17038
QY 1693 lnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyL 1713
D 17037 -----CCGCTCCGCGAAGCGCGGCTCTGCTCCAGCAGCGCCCGGCC 16995
QY 1713 euSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleA 1733
D 16994 CATCCCGCGCACTGGCTCCCTT-----GGCCGGGGAACACGAACACCACTT 16947
QY 1733 spLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaT 1753
D 16946 GCCTCGCGCAGCGCGCTACCCCGC-----ACACCGCCCGGTGGCGCGACCTTGCGA 16893
QY 1753 hrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHis 1773
D 16892 CAGCG-----CCCGCAGCGCTCTCCCTCGCTCGGACAGCTGCG 16854
QY 1773 erSerSerProLeu-----SerProGlyGlyProThrHisLeuThrL 1787
D 16853 CGCAAGCAGCATCGCGTGAGCGAAGTGGTCCGCTGCGCGCGCTGCTCTCCACCAC 16794
QY 1787 ysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgA 1807
D 16793 GTCCGACCATGCACTCCCGCTGCTCCGAGCCACTTCGCGCAGCGCTCCGCTCGCGC 16734
QY 1807 spArgGluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaProIle 1827
D 16733 CCCCACGAGCGCTGCTGCTCGCCCGACACACAGCGGATCGCAACACCTTCGTCGC 16674
QY 1827 rpArgProGlyThrGluGlnSerSerSerGlySerSerGlySerGlyGlyGlyGly- 1846
D 16673 C-AACGGCTCGCATTTGGCTCCACCGCTCGCGCGCGCTCCACCGCGCGCTCTTCGA 16615
QY 1847 -----SerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerPro- 1862
D 16614 GGATGACATCGCGGT--CGTCCCGCTGATCCCGAAGCAGACAGCGCGCGCGCGCGCC 16556
QY 1863 -----IleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeu-His 1879
D 16555 CGCGCGTTGCGCGCCACGACGCGGCTCTTGACGACACAGACAGCGCTCCCTCCAC 16496
QY 1880 ---AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu 1898
D 16495 CCAATGTGCGGGCTCGGCTGCTCGCGTGCGAGCTCGCGGACAGCACCTCGTCTGCATC 16436
QY 1899 ArgSerThrSerThrSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
D 16435 GACAGACCACTTTGAGCACACCGGCGACCGCGCGCA-----GGCCCGCGG- 16388
QY 1919 CysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeu 1938
D 16387 TGTCCA---GTTTCGACTTCGACGACCAAGGTACAGCGCGCTCGGCTTCGCGCCAG 16331
QY 1939 ProLysGluAlaProArgValAlaArgProGluArg----- 1950
D 16330 GCCCGAACACCTCCACGAGCGCTCCGCGCTCGATCGGCTCGCGCAGGTTCGTCGCGTCC 16271
QY 1951 -----ProArgAlaAspThrGly-----HisAlaPheLeu 1960
D 16270 CATGCGGCTCCACGCGCTCGATGTCTCGCGCGACAGCGCGCACAGAGAGCGCTGTT 16211
QY 1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
D 16210 GGATCACCGCTGCTGGCGAGGCGCTTGGCGCGCTCAACCTCGCTGCGCGACCGCTCT 16151
```

```
Db 19121 CGACCGCCACCGCGCGCGCGCTCGCGCCCTGCTCGAGCGCGCGCAACCAATGCTTCGA 19062
Qy 1082 -----HisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg 1098
Db 19061 GATCGGCAATCGCTTCGCGCCCTCAGGCGTT-----CGGCGAGCTCGCTCGTC 19014
Qy 1099 ProProThrIleSerAenProProProLeuIleSerSerAlaLysHisProSerValLeu 1118
Db 19013 CTTCGGCTGACCGACGACACCA-----GCGAGCCGCGCTCCAGGGGGCTG 18966
Qy 1119 GluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSer 1138
Db 19065 CCGCGA-----GCTCCACAGGCTGGAAGCTCACTGATACATATGTTGCCACAG 18915
Qy 1139 GluHisAlaLysAlaProValGlyProValThrMetGlyLeu----- 1152
Db 18914 CGTGAACGGGCTCCGCGAGCTGCTCGGCGCTCGCGCGCGCAGATGTAGAGCGCGGATGC 18855
Qy 1153 ---ProLeuProMetAspProLysLeuAlaProPheSer----- 1165
Db 18854 TCGCCACGGCT-----GCCCGTGGAGTCTGCGACTTGCAGCGAAG 18813
Qy 1166 -----GlyValLys-----GlnGluGlnLeuSerPro 1174
Db 18812 CGGTGCTCTGCTGCGCCTCGTCTGGAGCTCTACCGCGACCGCGAAGCTCGCTCGCC 18753
Qy 1175 ArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerVal 1194
Db 18752 CCGTCGGGTGAGCGCCAGCT-----CCGACCGAGCAACCGGACGAGCA 18708
Qy 1195 LeuArgGlyThrAlaLeuGlySerValProGlySerIleThrLysGlyIleProSer 1214
Db 18707 CTTGTTG-----CCCTGGGCTCTCGGAGAGCGCGGACCATCG 18666
Qy 1215 -----ThrArgValProSerAspSerAlaIleThr---TyrArgGlySerIleThrHis 1231
Db 18665 TGTGAGCGCGGCTCATCAAC-GCGGGGTGACGCGCATACGCTCCGCGCTGCTCCTTC 18607
Qy 1232 GlyThrProAlaSerValLeuTyrLysGlyThrIleThrArg-----Ile-I1 1247
Db 18606 TCGGTCCCC-----GGCAGCACCAACCGGACCGAAGTAGGTGTGTGCT 18565
Qy 1247 eGlyGluAspSerProSerArgLeuAspArg---GlyArgGluAspSerLeuProLysG1 1266
Db 18564 TGGCGGACAGCTCCAGAGCCCTTGAACCGCGGCGCT----- 18525
Qy 1266 yHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyMetSerVa 1286
Db 18524 -----AGTGGAGCGCGCTCACGACGCGCT----- 18498
Qy 1286 lThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro-----ProHi 1302
Db 18497 -----CGTAAAGCCCGGACAGGTCCACCTCCTCCGCGCCCGC 18460
Qy 1302 sGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaI1 1322
Db 18459 ACAGGCCACGCTGCACACTGTCGAGCGGACAGGATGCGGGGGCTCGCTCGGTCAACA 18400
Qy 1322 eSer-----SerAlaSerIleGluGlyLeuMetGlyArgAlaIlePr 1336
Db 18399 CTCGGCTGGGTGCTGCACCCAGCGCGCTCTCCGGGCGCTGCTCAGGCTGGGTGTACA 18340
Qy 1336 oProGlu-----ArgHisSerProHisIleLeuLysGlu----- 1347
Db 18339 GCCCCACCTCGCGCGCGCGCGCTCGCGCGCTCCGATCATCATCTGACGCGCGCGC 18280
Qy 1348 -----GlnHisIleArgGlySer-----IleThrG1 1357
Db 18279 CGCCTCGGCGAGCACCGCGCTCGCGCGCTCGCGAGGTGAGTCCGACAGCGCGGTGCCA 18220
Qy 1357 nGlyIleProArgSerTyrValGluAlaGlnGluAspTyrIleuArgGlu-----Al 1375
```

QY 510 -----GlnProMetProArgSerSerGlnG1 518
Db 21196 AGCACGCCGTGTCACCGTCATCGCTGGCGCTGACAGCCCGCAG-----CGCGTAGGCCACCC 21140
QY 518 uGluLysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 538
Db 21139 GGCCCGAGAGGATGCTCGCGCGCTCCCGATGCCCGGTACCGTCGTCAGATCCTTCAGAT 21080
QY 538 lGluAsnAspLysGluAspLeuLys---GluLysThrAsp---AspThrSerGlyG1 556
Db 21079 CGACAGTATGACAGCAGCAGCAGCTAGTCGAGCCCATCGAACCCAGGTACACTCCGGTGGCG 21020
QY 556 uAspAsnAspGluLysGluLysValaValaSerLys----- 567
Db 21019 TCCCGCTCAGCGCGGCGGCGCACGCGCGCAGCTCCAATGCTCCACGCGCTCTCCA 20960
QY 568 -----GlyArgLysThrAlaAsnSerG1 575
Db 20959 GCACAGCCGATCTCGGATCCATCGACTGCGCCTCGCGCGCGATA-TCCCGAAGAAG 20901
QY 575 nGlyArgLysGlyArgLysThrArgSerMetAlaAsnGluAlaAsnSerGluGluAl 595
Db 20900 TTCGCTCGAACAGTTCGATGTCGGAACAAACCCACTTCGCGCGGTAGCTCTTGCT 20841
QY 595 al1erThrProGlnGlnSer-----AlaGluLeuAlaSerMetG1 608
Db 20840 GCGGCTCCGATCAGGTCGTAGACTCGAACCCGCTCCAGCGCGCTGGGAGCGCTTCG 20781
QY 608 uLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLe 628
Db 20780 ATCGCGCTCTCCCTCGGCAAGACGCGCAGTAGTCGTCGCGCGCTCGCAACGCTCCC 20721
QY 628 uGluHisGly-ArgAsnTrpSerAlaAlaLeuAlaArgMetValGlySerLysThrValSerG 648
Db 20720 GGCAGCGCGCACCCATCGACAGTCGCGATCGCTCTCTTCTTCTTCTGTCACACGCGC 20661
QY 648 lncysLysAsnPheThrPheAsnTyrLysLysArgGlnAsnLeu----- 662
Db 20660 ACCCGCGCGCGCTTCTTCACTCGACGCGCGCGAGAACCGCTGCTTCAGCAGCAGCTCT 20601
QY 663 -----AspGluLeuGlnGlnHisLysLeuLysM 673
Db 20600 GCGATCGCGCGCGCTCGGGTAGTCAAGACACGCTCGCGCGCAAGGACGCTCTCGGTA 20541
QY 673 etGluLysGluArgAsnAlaArgLysLysLysLys----- 685
Db 20540 CCGGACGTGAGCGGTTCGCGACGCGCCACCGCCATCAGCGAGTCGAGCGCCCGCTCTCTG 20481
QY 686 -----AlaProAlaAlaSerGluGluAlaAlaPheProp 698
Db 20480 GCACCTGATCCGTGCGCACAGCTCGCGCGCTGAGCGCTCCACGCGCGCAACCTCGC 20421
QY 698 roValValGluAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysMetV 718
Db 20420 CCGGACACGCTCGAGCGCATTCAGCGCGCTCCGCGCGCGCGCGAGCGCGCT 20361
QY 718 alGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerG 738
Db 20360 CCGGAGCGCGCGAGCGCTCTCGCTCTGTCGAGGAGCGCTTCGCGAAG-----CTCG 20307
QY 738 lyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSer-ProHisThrGlu 757
Db 20306 GCGCAACAGCGCACGAAACAGCGCGCGAGCTCGCGCTGGAGCTCCAGCGCCAGCTGCA 20247
QY 758 AlaAlaLysAspThrGly----- 763
Db 20246 GCTCGCGAGATCCAGGTGCACTGGGACCAAGGCTCGCTTCAGGCGGTGAGAGCGCGCGCT 20187
QY 764 -----GlnAsnGlyProLysProProAlaThrLeu----- 773
Db 20186 CCAGCAGAGGAGCGCCCTCTTCCACCGGATCGCGCAACCCATTCGCGCTCATCGCGG 20127

773 ----- 773
20126 ACAGCTCGGCCTCGCCAGGTGCGTGTATCCCAAGCGCCGCTGTGTCCAGAACCC 20067
774 -----Gly 774
20066 AGCCAGGCTCGTCGCGCGAGCCCGCGCGCGGTGCGCGCGAGCGGCGTCCAGGA 20007
775 AlaAspGlyPro-----ProGlyProPro-----ThrProProArg 787
20006 ACGTGTTCCGCGCGGTAGTTGCTCTGCGCGCGCAAAACGTACCGCGCGCGACG 19947
788 ArgThrSerArgAlaProLysGluProThrProAlaSerGluAlaThrGlyAlaProThr 807
19946 AGAACAGCAGCAACGCGCGAGATCCAGTCCGCGTCTGCTGTCAGCAGCGCGCC 19887
808 ProProAlaProProSerPro-----SerAlaProProProValProLys 824
19886 CGTCCACCTTCGCGCGCACCCCGCGAGGCGCTCGCGCGTCTGGGCGGTGAGCAGCG 19827
825 GluLysGluGluGluThrAlaAlaProProValGluGluGluGluGln--- 843
19826 CGTGTGCGAGCAGCGCGCGCGTGCAGCACCGCGCTCAGCGAGCGCGCGCTCGATGC 19767
844 LysProProAlaAlaGluLeuAlaValAspThrGlyLysAlaGluGluProValLys 863
19766 CGGCCAGCACGCGCGAGCTCTCTCCGCTTCGACACGTCGACGAGCAGCAGCCGTCACCG 19707
864 SerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
19706 TCTCGCGCGAGCTCTCGAGTGCAGCAGCTCGCGCGCGCGCGCGCGCTCCAGCC 19647
884 GluAlaThrAlaGluLysAlaLeuLysAlaGluLysGluGlySerGlyArgAla 903
19646 CGC-----GACGCGCGTCAGCACAGTCCCTCACCCCT--- 19611
904 ThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAla 923
19610 -----GCGCGCGCACAGTTCGCGCGATCGCTGACCCAGCTCCCTGTCCCGCGG 19557
924 AspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeu-----Leu 938
19556 TCACGAGCAGCTGCGCGCGAGGTTCAGTCTCGCGGCTCTCGGTGAGTCTTCGCTG 19497
939 SerPro---ArgProSerLeuLeuThrProThrGlyAspProArgAlaAlaSerPro 957
19496 CCGCTCGCGCGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19437
958 GlnLysProLeuAspLysGlnLysGlnLysGlnLysGlnLysAlaAlaAlaProLysGln 977
19436 GCTCCGCG-----CGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19398
978 ValThrLysValHisGluProProArgGluAspAla----- 989
19397 GTTCCAGGCTGCTGCTCCCATCATCATCAAGCGAGCGCGCGCGCGCGCGCGCGCG 19338
990 AlaProThrLysProAlaProAla-----ProProProProGlnAsnLeuGlnPro 1007
19337 GCGCGCGCGCAACAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19278
1008 GluSerAspAlaProGlnProGlySerSerProArgGlyLysSerArgSerProAla 1027
19277 GCG 19218
1028 ProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPro 1047
19217 CCAGCAGCGCGCTTCGAGCGCGAGCG-----CTGCTGCTGCTGCT 19173
1048 CysTrpThrSerGlyLeuProPheProValProArgGluValLysAlaSerPro 1067
19172 -----CGTGGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19122
1068 HisAlaProAspProSerAlaPheSerTyrAlaProGly----- 1081

ketoreductase domain and acyl carrier protein domain. A host cell comprising a PKS ORF (open reading frame) which encodes one or more PKS domains is useful for producing polyketide synthases from which polyketides can be produced. The host cells are useful for constructing a library, where each individual colony of the library represents a colony with the ability to produce a particular PKS synthase and ultimately a particular polyketide. The polyketides produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity. Colonies in the library are also induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of candidate polyketides which can be screened for binding to desired targets such as receptors, signaling proteins, etc. The present sequence represents the DNA sequence of cosmid pKOS28-26 which encodes one or more domains of *S. cellulosum* PKS. Note: The present sequence is said to encode the functional domains of *S. cellulosum* PKS which correspond to domains or domain subsets of the amino acid sequences of ORP1 (AAU10700) and ORP2 (AAU10701). (Updated on 11-SEP-2003 to standardize QS field)

Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 0 U; 6 Other; XX

| | |
|------------------------|----------|
| Alignment Scores: | |
| Pred. No.: | 2.08e-13 |
| Score: | 726.50 |
| Length: | 33529 |
| Matches: | 674 |
| Conservative: | 314 |
| Mismatches: | 1199 |
| Best Local Similarity: | 22.08% |
| Query Match: | 5.50% |
| Indels: | 873 |
| Gaps: | 137 |
| DB: | 5 |

US-09-522-753-5 (1-2517) x AAS17367 (1-33529)

| | | | |
|----|-------|---|-------|
| Qy | 17 | ProArgTyProPro---HisSerLeuSerTyPrProValGlnIleAlaArgThrHisThr | 35 |
| Db | 22686 | CCGAGCGCTCCACCGACACCATCGCTCCCGCGCTCGCACCTTGCA-----TCAGCC | 22633 |
| Qy | 36 | AspValGlyLeuLeuGluTyGlnHisHisSerArgAspTyAlaSerHisLeuSerPro | 55 |
| Db | 22632 | GA CCTCGGG-----CGCACACTAGCTTCGCTGGCTCCGCAA | 22597 |
| Qy | 56 | GlySerIleIleGlnProGlnArgArgProGlnLeuSerGluPheGlnProdgly | 75 |
| Db | 22596 | GGCTCAGACACACCGGCACATGCGACGCGCTCAGCTCTCTATCCAGT---GCCCCACGA | 22540 |
| Qy | 76 | AsnGluArgSerGln-----GluLeuHis----- | 83 |
| Db | 22539 | GCGCAGGGCTCAGCCCCACGATCCCATGACGGTACAGCGCCACTTCACGGCGAAA | 22480 |
| Qy | 84 | -----LeuArgPro-----GluSerHisSerTyLeuProGluLeuGly | 96 |
| Db | 22479 | CAGCCCGGGTGCGTGTACTCCGTCGCTCGACGAACGCTCTCTCTCGCTCGCCGCGGC | 22420 |
| Qy | 97 | LySer-----GluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeu | 113 |
| Db | 22419 | CGGGAACACACCTCTCTCAACCCACGGTCGAGATACGGCTCCAGCGCTCGCACACCTC | 22360 |
| Qy | 114 | ProAsp-----ProLeuLeuArgProSerPro | 122 |
| Db | 22359 | GTCGAACCGCGCACGGAAACACGGGGTACACTTCGTAAGAGCTCTTCCCATCCGAGCGC | 22300 |
| Qy | 123 | LeuLeuAlaThrGlyGlnProAlaGlySerGluAAspLeuThrLysAspArgSerLeuThr | 142 |
| Db | 22299 | CTGGCTGCCCTG-----CCCGGTGAACAGCAC | 22273 |
| Qy | 143 | GlyLysLeuGluProValSerProProSerProHisThrAspProGluLeuGluLeu | 162 |
| Db | 22272 | CGCAAGCTT---CCCGCTCGCTTCGCGTCCGCTCAC-----CACCGCGCGCAT | 22225 |
| Qy | 163 | ValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGlu | 182 |
| Db | 22224 | CGGCGCGCCCGACGACAGCGCGGAGAGCCCTCCACAGCTCCCGCAGCGCTCGCCGCAAG | 22165 |
| Qy | 183 | IleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeuGluGlu | 202 |

Db 566 GAACAAAAATCTGCCAGCCTTATGATCAGCTCATGGAGCGCATGGAGAAAAAAGTGGAC 625
QY 315 ArgileGluAsnProArgArgAlaLysGluSerLysValArg 330
|||
Db 626 AGATAGAAATATCTCGAGGAAAGCTTAAGAAAGCAACACAGG 673
RESULT 19
ABK44934
ID ABK44934 standard; cDNA; 520 BP.
XX AC ABK44934;
XX AC
XX 05-JUN-2002 (first entry)
XX DE cDNA encoding colon tumour protein, SEQ ID NO 485.
XX KW Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;
XX KW gene; ss.
XX OS Homo sapiens.
XX PN W0200212328-A2.
XX XX
XX PD 14-FEB-2002.
XX XX
XX PF 31-JUL-2001; 2001WO-US024218.
XX XX
XX PR 03-AUG-2000; 2000US-0223283P.
XX PR 28-MAR-2001; 2001US-0279763P.
XX PR 29-JUN-2001; 2001US-0302051P.
XX XX
XX PA (CORI-) CORIXA CORP.
XX XX
XX PI King GE, Meagher MJ, Xu J, Secretist H;
XX XX
XX DR WPI; 2002-241739/29.
XX XX
XX PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
XX PT for diagnosing, preventing, and treating colon cancer, and as markers for
XX PT the progression of cancer.
XX XX
XX PS Claim 1; SEQ ID NO 485; 147pp; English.
XX XX
XX CC The invention relates to polynucleotides encoding colon tumour proteins.
XX CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
XX CC compositions, such as vaccines, for the diagnosis, prevention, and
XX CC treatment of colon cancer. Polynucleotide sequences may be used as
XX CC hybridisation probes or primers, and in the design and preparation of
XX CC ribozyme molecules for inhibiting expression of tumour polypeptides and
XX CC proteins in tumour cells. The compositions are useful for stimulating an
XX CC immune response against cancer, particularly for the immunotherapy of
XX CC colon cancer, and as markers for the progression of cancer. ABK44450-
XX CC ABK46237 represent coding sequences of human colon tumour proteins of the
XX CC invention. Note: With the exception of SEQ ID No 1 and 2, the sequence
XX CC data for this patent did not form part of the printed specification but
XX CC was supplied by the European Patent Office
XX XX
SQ Sequence 520 BP; 199 A; 80 C; 128 G; 113 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.07e-15 Length: 520
Score: 746.50 Matches: 141
Percent Similarity: 91.95% Conservative: 19
Best Local Similarity: 81.03% Mismatches: 13
Query Match: 5.65% Indels: 1
DB: 6 Gaps: 1
US-09-522-753-5 (1-2517) x ABK44934 (1-520)
QY 285 LysArgArgAsnHisAlaArgLysGluTrpLysGluLysPheCysGlnArgTyrAspGln 304
|||
Db 1' AAGAAGAAGAAATCATGCAAGAAACAAAGGGAACAAAAATCTGCCAGCGTTATGATCAG 60

QY 305 LeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAla 324
|||
Db 61 CTATCGAGCGCATGGGAGAAAAAGTGGACAGAAATAGAAATATCTCGAGGAAAGCT 120
|||
QY 325 LysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArg 344
|||
Db 121 AAGAAGCAACAAACAAAGGAATACTATGAAAGCAGTTTCCAGAAATTCGAAACCAAGA 180
|||
QY 345 GluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSer 364
|||
Db 181 GAACAGCAAGAAAGATTTCAG---CCAGTTGGCAGAGGGAGCTGGTCTTTCCAGCCACC 237
|||
QY 365 AlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsn 384
|||
Db 238 ATTGTAGGAGTGAGCATGAGATTTCTGAAATTTATTGATGGGCTCTCTGAGCAGGAGAA 297
|||
QY 385 LeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAlaAspGln 404
|||
Db 298 AATGAGAAACAAATGCGCAGCTCTCTGTGATTCACCTATGATGTTTGTATGAGCAACAA 357
|||
QY 405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys 424
|||
Db 358 AGACGAGTCAAGTTTCATTAAACATGATGGCTTATGAGGACCTATGAAAGTGATATAA 417
|||
QY 425 AspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPhe 444
|||
Db 418 GATAGGCAGTTTATGAATGTTTGGACTGACCATGAAAGAGAGATCTTTAAGGACAAGTTT 477
|||
QY 445 MetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeu 458
|||
Db 478 ATCCAGCATCCAAAAAACTTTGGACTTAATTGCATCATACTTTG 519
|||
RESULT 20
AAS17367/c
ID AAS17367 standard; DNA; 33529 BP.
XX AC AAS17367;
XX DT 11-SEP-2003 (revised)
XX DT 12-MAR-2002 (first entry)
XX XX
XX DE DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX XX
XX KW Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
XX KW acyl transferase domain; dehydratase domain; ketoreductase domain;
XX KW acyl carrier protein domain; pKOS28-26; ds.
XX OS Polyangium cellulosum.
XX XX
XX PN US6280999-B1.
XX PD 28-AUG-2001.
XX XX
XX PF 31-AUG-1998; 98US-00144085.
XX PR 22-JAN-1998; 98US-00010809.
XX XX
XX PA (KOSA-) KOSAN BIOSCIENCE.
XX XX
XX PI Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX XX
XX DR WPI; 2001-606536/69.
XX XX
XX PT Novel purified, isolated DNA molecule from Sorangium cellulosum having
XX PT polyketide open reading frame encoding modules with one or more domains
XX PT such as ketosynthase, acyl transferase and acyl carrier protein domains.
XX PS
XX PS Claim 4; Fig 1; 72pp; English.
XX XX
XX CC The present invention relates to the isolation of novel Sorangium
XX CC cellulosum polyketide synthases (PKS), and the polynucleotide sequences
XX CC encoding them. The polyketide synthases include catalytic domains such as
XX CC ketosynthase domain, acyl transferase domain, dehydratase domain,

| | | | |
|-----|----|---|-----|
| 475 | Db | GAGATTCTCGAAATTAATTGATGGGCTCTCGAGCAGGAGAAATATGAGAAACAATATCGCG | 416 |
| 391 | Qy | GlnLeuAlaValIleProProMetLeuTyrzAspAlaAspGlnGlnArgIleLysPheIle | 410 |
| 415 | Db | CAGCTCTCTGTGATTCACACCTATGATGTTTCATCGAGAACAAAGACGAGTCAAGTTCAIT | 356 |
| 411 | Qy | AsnMetAsnGlyLeuMetAlaAspProMetIlySValTyrLysAspArgGlnValMetAsn | 430 |
| 355 | Db | AACATGAATGGGCTTATGGAGACCCATGAAGAGTGATAAGATAGTAGCAGTTTATGAAT | 296 |
| 431 | Qy | MetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsn | 450 |
| 295 | Db | GTTTGGACTGACCATGAAAGAGAGATCTTTAAAGACAGATTTATCCAGCATCCAAAAAC | 236 |
| 451 | Qy | PheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyr | 470 |
| 235 | Db | TTTGACTAATTTGCATCATCTTGGAGAGAGAGATGTTCTCGATTGTGTTTTGTATTATAC | 176 |
| 471 | Qy | TyrLeuThrLysLYSAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArg | 489 |
| 175 | Db | TATTTAAACCAAGAAAAAATGAGAAATTATAAGCCCTCGTCAGAGGAATATGGAAACGC | 116 |
| 490 | Qy | ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln | 509 |
| 115 | Db | AGAGGCAGAAACACAGCAA----- | 98 |
| 510 | Qy | GlnProMetProArgSerSerGlnGluLysAspGluLysGluLysGluLysGluAla | 529 |
| 97 | Db | -----ATTGCTCGACCTTCGCAGAGAGAAAAAGTAGAGAAAAAGAGAGGATAAGCA | 44 |
| 530 | Qy | GluLysGluGluGluLysProGluValGluAsnAspLysGlu | 543 |
| 43 | Db | GAATAAACAGAAAAAAGAGAGAGAAAAAGAGAAAGATGAAGAG | 2 |

RESULT 18

| | | | |
|----|-----|--|-----|
| Qy | 116 | ProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeu | 133 |
| Db | 38 | CTTTTAGTCACCCCGCTGCCA-----GAAGGGCTGAGGGCT---TCTGCAGATGCT | 85 |
| Qy | 136 | ThrLysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProProSerProPro | 154 |
| Db | 86 | AGAAGGATCCAGCATTCGGAGGCAACAACATGAAGCTCCATCTCTCCAATTTCGGGGCAA | 145 |
| Qy | 155 | HisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGln | 174 |
| Db | 146 | CCATGTGGAGATGATCAAAATGCTTCACCTTCAAAACTCTCAAGAGAGAGTTAATACAG | 205 |
| Qy | 175 | AsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLys | 194 |
| Db | 206 | AGTATGGATCGTGTAGATCAGAAATTCGAAAGATGAAACAGCAGATCCTTAACCTGAA | 265 |
| Qy | 195 | LysLysGlnGlnGlnLeuGluGluAlaAlaLysProProGluProGluLysProVal | 214 |
| Db | 266 | AAGAAACAACACAGCTTGAAGAAGAGGAGCTAAACCTCTCTGAGCTTGAGAAGCCCGTG | 325 |
| Qy | 215 | SerProProIleGluSerLysHisArgSerLeuValGlnIleIleIleTyrAspGluAsn | 234 |
| Db | 326 | TCCCTCTCCTCTGTGGAGCAGAAACACCGCAGTATTGTCCAAATTTATTATGATGAGAAT | 385 |
| Qy | 235 | ArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu | 254 |
| Db | 386 | CGGAAAAAAGCAGAGAAGCTCATAAATTTTGAAGGCTCTTGCGCCAAAAGTTGCAACTG | 445 |
| Qy | 255 | ProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGln | 274 |
| Db | 446 | CCACTGTATACCCAGCCATCAGATACCAAGGTGTACCATGAGAAACATCAAGACCAACAG | 505 |
| Qy | 275 | AlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyr | 294 |
| Db | 506 | GTGATGAGGAAAAAACTCATTTTATTTTAAAAAGAGAAATCATGCAAGAAAAACAAAGG | 565 |
| Qy | 295 | LysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu | 314 |

RESUL16

ACAS57523 standard; cDNA; 673 BP.

ACAS57523;

10-JUN-2003 (first entry)

Human adipocyte Selected Interacting domain, SID, cDNA #610.

Human; ss; gene; prey; adipocyte; SID; selected interacting domain; anorectic; antidiabetic; protein-protein interaction; diabetes; yeast 2-hybrid assay; metabolic disorder; obesity.

Homo sapiens.

WO200286122-A2.

31-OCT-2002.

14-MAR-2002; 2002WO-EP003768.

14-MAR-2001; 2001US-0275734P.

(HYBR-) HYBRIGENICS.

Legrain P, Daviet L;

WPI; 2003-103412/09.

P-PSDB; ABU70979.

New complex between two interacting proteins in adipocyte cells, useful for identifying selected interacting domains that modulate protein interactions, or for preventing or treating metabolic disorders such as obesity or diabetes.

Claim 7; Page 308-309; 382pp; English.

The invention relates to a complex between two interacting proteins in

DE Human adipocyte Selected Interacting domain, SID, cDNA #611.
XX Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
KW anorectic; antidiabetic; protein-protein interaction; diabetes;
XX yeast 2-hybrid assay; metabolic disorder; obesity.
OS Homo sapiens.
XX WO200286122-A2.
XX 31-OCT-2002.
XX PF 14-MAR-2002; 2002WO-EP003768.
XX PR 14-MAR-2001; 2001US-0275734P.
XX (HYBR-) HYBRIGENICS.
XX Legrain P, Daviet L;
XX WPI; 2003-103412/09.
XX P-PSDB; ABUT0980.
XX New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes.
XX Claim 7; Page 309; 382pp; English.
XX The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC polynucleotide encoding a polypeptide in the adipocyte cells, a
CC recombinant host cell expressing at least one of the interacting
CC polypeptides of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
CC sequences given in the specification (including its fragment or variant),
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC given in the specification (including its fragment or variant), a vector
CC comprising the SID (RTM) polynucleotide, a recombinant host cell
CC comprising the vector, a protein chip comprising the polypeptides and a
CC record comprising all or part of the data, listed in the specification.
CC The complex, polypeptides, polynucleotides and compounds are useful for
CC preventing or treating metabolic disorders such as obesity or diabetes.
CC The polynucleotides are useful as probes or primers. The complex is
CC particularly useful for identifying selected interacting domains (SID
CC (RTM)) for screening drugs that modulate the protein interaction, thus
CC exhibiting the therapeutic effect. The present sequence encodes a SID
XX (prey) protein of the invention
XX
SQ Sequence 555 BP; 151 A; 162 C; 167 G; 75 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.33e-22 Length: 555
Score: 960.00 Matches: 184
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.26% Indels: 0
DB: 7 Gaps:

US-09-522-753-5 (1-2517) x ACA57524 (1-555)

QY 138 AspArgSerLeuThrGlyLysLeuGluProValSerProSerProHisThrAsp 157
DB 2 GACGTAGCTGACGGGACAGTGGAAACCGGTGTCTCCCGCCAGCCCGCCACCTGAC 61

QY 158 ProGluLeuGluLeuValProProArgLeuSerLysGluLeuLeuLeuGlnAsnMetAsp 177
DB 62 CCTGAGCTGAGCTGTGTGCGCCACCGCTGTCCAGGAGGAGCTGATCCAGAACATGGAC 121

QY 178 ArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGln 197
DB 122 CGGTGGACCGAGAGATCACCATGGTAGACGACGATCTCTAAGCTGAAGAAGACAG 181
QY 198 GlnGlnLeuGluGluAlaLysProGluProGluLysProValSerProPro 217
DB 182 CAACAGCTGGAGGAGGAGCTGCAAGCCGCCGAGCTGAGAAGCCCGTGTACCCGCCG 241
QY 218 ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 237
DB 242 CCATCGAGTCGAAGCAGCCGACCTGGTGCAGATCATCTACCGAGAACCCGGAAGAG 301
QY 238 AlaGluAlaAlaHisArgIleLeuGluGlyProGlnValGluLeuProLeuTyr 257
DB 302 GCTGAAGCTGCACATCGGATTCGGAAGGCTGGGCCCCAGGTGGAGCTGCCGTGTAC 361
QY 258 AsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArg 277
DB 362 AACGAGCCCTCCGACACCCGCGCAGTATCATGAGAACATCAAAATAAACACGAGCGATCGG 421
QY 278 LysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLys 297
DB 422 AAGAAGCTAATCTTGTACTTCAAGAGGAGGAATCAGCTCGAAACAATGGAAGCAGAAG 481
QY 298 PheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGlu 317
DB 482 TTCTGCAGCGCTATGACAGCTCATGGAGCCCTTGGAAAAAAGGTGGAGCGCATCGAA 541
QY 318 AsnAsnProArg 321
DB 542 AACACCCCGCGC 553

RESULT 16
ADE76358/c
ID ADE76358 standard; DNA; 718 BP.
XX ADE76358;
XX AC ADE76358;
XX DT 29-JAN-2004 (first entry)
XX DE Human BSK-23 DNA.
XX KW monocyte; macrophage; gene expression profile; rheumatoid arthritis;
KW chronic inflammatory disease; bacteria-induced inflammation;
KW arteriosclerosis; tumour; organ; tissue transplant; sepsis;
XX molecular classification; human; BSK; ss.
XX OS Homo sapiens.
XX PN EP1310567-A2.
XX PD 14-MAY-2003.
XX PF 02-OCT-2002; 2002EP-00090348.
XX PR 09-NOV-2001; 2001DE-01055600.
XX PA (OLIG-) OLIGENE GMBH.
XX PI Stuhlmueller B, Haeupl T;
XX DR WPI; 2003-443090/42.
XX PT Device for diagnosis, prognosis and monitoring therapy of e.g. rheumatoid
PT arthritis, comprises immobilized gene sequences from monocyte-macrophage
XX cells.
XX PS Claim 29; SEQ ID NO 171; 180pp; German.
XX CC The invention relates to a novel device for diagnosis, prognosis and
CC monitoring of therapy having, on its surface, sequences of some or all of
CC specified monocyte-macrophage genes and also of other genes or RNA

RESULT 14

AAA02670
ID AAA02670 standard; cDNA; 872 BP.

XX
AC AAA02670;

XX
DT 19-MAY-2000 (first entry)

XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2661.

XX
KW Human; colon cancer; tumour; diagnosis; gene expression product; probe;

XX
KW detection; cancerous state; metastasis; identification; breast cancer;

XX
KW oestrogen receptor-positive breast cancer; therapy;

XX
KW oestrogen receptor-negative breast cancer; lung cancer; ss.

XX
OS Homo sapiens.

XX
PN WO958675-A2.

XX
PD 18-NOV-1999.

XX
PF 13-MAY-1999; 99WO-US010602.

XX
PR 14-MAY-1998; 98US-0085426P.

XX
PR 15-MAY-1998; 98US-0085537P.

XX
PR 21-OCT-1998; 98US-0085696P.

XX
PR 27-OCT-1998; 98US-0105234P.

XX
PR 27-OCT-1998; 98US-0105877P.

XX
PA (CHIR) CHIRON CORP.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

XX
PI Reinhard C, Giese K, Randazzo P, Kennedy GC, Pot D, Kassam A;

XX
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

XX
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

XX
XX WPI; 2000-126369/11.

XX
XX Polynucleotide library used to determine cancerous states of mammalian

XX
XX cells.

XX
PS Claim 1; Page 1079; 1097pp; English.

XX
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA

XX
CC libraries constructed from human colon cancer cell lines. The present

XX
CC invention also describes a method of detecting differentially expressed

XX
CC genes correlated with a cancerous state of a mammalian cell, comprising

XX
CC detecting at least one differentially expressed gene product in a test

XX
CC sample derived from a cell suspected of being cancerous, where detection

XX
CC of the differentially expressed gene product is correlated with a

XX
CC cancerous state of the cell from which the test sample was derived. The

XX
CC polynucleotide sequences can be used in a method for detecting

XX
CC differentially expressed genes correlated with a cancerous state of a

XX
CC mammalian cell. The polynucleotides can also be used as probes for

XX
CC detecting and mapping related genes. They can be used in diagnosis and

XX
CC prognosis of diseases and disorders (e.g. identification of pre-

XX
CC metastatic or metastatic cancerous states, stages of cancer, or

XX
CC responsiveness of cancer to therapy). This is particularly for breast

XX
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-

XX
CC negative breast cancer, lung cancer, and colon cancer

XX
XX

XX
SQ Sequence 872 BP; 345 A; 133 C; 203 G; 185 T; 0 U; 6 Other;

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 10 GCTGTGCAAAATTTTGAAGGTCTTGCCCAAAAGTTGAAGTCCACTGTATTAACAGCCA 69
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 70 TCAGATACCAAGGTGTACCATGAGAACATCAAGACAAACAGGTGATGAGGAAAAAATC 129
QY 281 IleLeuTyrPhelYsArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
DB 130 ATTTTATTTTAAAGAGAAATCATGCAAGAAACAAAGGAAACAAAAATCTGCCAG 189
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
DB 190 CGTTATGATCAGCTCATGGAGCATGGGAGAAAAAAGTGCACAGAAATAGAAAAATATCT 249
QY 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluI 340
DB 250 CGGAGGAAAGCTAAAGAAAGCAAAACMAGGGAATACTATTAAAAAGCAGTTTCCAGAA 309
QY 340 eArgLysGlnArgGluLeuGlnArgMetGlnSerArgValGlyGlnArgGlySerG 360
DB 310 TCGAAAAACAAAGAGACACAGAAAGATTTCAG---CGAGTTGGCAGAGGGAGCTGG 366
QY 360 YLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSe 380
DB 367 TCTTTTCAGCCACCATTGCTAGGAGTGCAGCATGAGATTCTGAAATTTATTGATGGGTCTC 426
QY 380 rGluGlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProMetLeuTyr 400
DB 427 TGAGCAGGAGAAATAATGAGAAACAAATGCGGAGCTCTC-GTGATTCCACCTATGATGTT 485
QY 400 rAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProme 420
DB 486 TGATCGCAACAAAGACGAGTCAAGTTCATTATCATGAATGGCTTATGGAGAGACCTAT 545
QY 420 tLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
DB 546 GAAAGTGATATAAGATAGGCAGTTTATGAATGTTTGGACTGTACCATGAAAGAGATCTT 605
QY 440 eArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluAr 460
DB 606 TAAGCAAGTTTATCCAGATCCAAAAAATTTGGACTAAATTCATCATCTACTTTGGAGAG 665
QY 460 GlyThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLy 480
DB 666 GAAGAGTGTCTCGATTGTTTGTATTACTATTAAACCAAGAAAAATGAGAAATTATAA 725
QY 480 sSerLeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnG 499
DB 726 AGCCCTCGTCAGAAAGGAATATGGGAAACGCAGAGGCAGAAACAGCAA----- 774
QY 499 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 519
DB 775 -----ATTGCTCGACCCCTCGCAAGAGA 797
QY 519 uLysAspGluLysGluLysGluAlaGluLysGluGluGluLysProGluValG 539
DB 798 AAAAGTAGAGAAAAAGAGAGAGATAAAGCAGAAAAAACAARAAAAAAGAGAGAAA 857
QY 539 uAsnAspLysGlu 543
DB 858 GAAAGATGAAGAG 870
RESULT 15
ACA57524
ID ACA57524 standard; cDNA; 555 BP.
XX
AC ACA57524;
XX
DT 10-JUN-2003 (first entry)
XX

Alignment Scores:
Pred. No.: 2,52e-25 Length: 872
Score: 1076.00 Matches: 216
Percent Similarity: 82.62% Conservative: 36
Best Local Similarity: 70.82% Mismatches: 34
Query Match: 8.14% Indels: 20
DB: 3 Gaps: 3

US-09-522-753-5 (1-2517) x AAA02670 (1-872)

| | | | |
|----|-------|--|-------|
| Db | 9146 | CACGCTCAA-----TCCAGATCGCGAACGGGAATCGTACTATCTGTCAGG | 9190 |
| Qy | 1900 | SerThrSerThrSer---SerProValArgProAlaAlaThrPheProProAlaThrHis | 1918 |
| Db | 9191 | CACACGGCGGTCCAGTCCAGAGGATACGCCCGGCAACTGAGCGCCAGAGTCTGATCG | 9250 |
| Qy | 1919 | CysProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeuLeu | 1938 |
| Db | 9251 | ATGCCA---TCATCAAGCATGAGATCAATCGTTCCA-----ATGATGCCACCGCTGGC- | 9300 |
| Qy | 1939 | ProLysGluAlaProArgValAlaArgProGluArgProAlaAspThrGlyHisAla | 1958 |
| Db | 9301 | CCGGGTGCG-----GAATTCGCGCGCCCATCGTTCGTTCATGCT | 9339 |
| Qy | 1959 | PheLeuAlaLysProProAlaArgSer-----GlyLeuGluProAlaSerSer | 1974 |
| Db | 9340 | CCGTTG-----CCGCCACGTGGTTCCGGATCGGGAGCGGCCCATACAGTCTCTCA | 9393 |
| Qy | 1975 | ProSerLysGlySerGluPro-----ArgProLeuValProProValSerGlyHis | 1991 |
| Db | 9394 | CCGCCCAATGTCTCCATCCGATGTACTCGCGGATCTACGCCAGCCGCTCGACGGCGGA | 9453 |
| Qy | 1992 | AlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspPro | 2011 |
| Db | 9454 | CGGGATCAATGCTTACGGCGGAGAACAAATGCCAAGCCC-----AGCTCGTGGGATCG | 9507 |
| Qy | 2012 | ProAla-----ProAlaSerAlaSerAsp | 2020 |
| Db | 9508 | CCCACTGATTAATATTGATTGTCAGAGCGCATCTCAGCGCGAGCGGACGACGTT | 9567 |
| Qy | 2021 | ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuLeuArgSer | 2040 |
| Db | 9568 | CGCCAGCAGCAGCAGCAGCAGCAAGCTCCGCTCCGTCGCAATCGTCGAGTCTAGATCC | 9627 |
| Qy | 2041 | LeuGlyTyrHisGlySerSerTySerProGluGlyValGluProValSerProValSer | 2060 |
| Db | 9628 | GTG-----CACGTGACGTAAAGCACCGACCTCG---CAAGCTGGCGGATCAGCGCCC | 9678 |
| Qy | 2061 | SerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer | 2080 |
| Db | 9679 | AGTCACACAGCAGATTCAACAGAGCATTACC-----TTTGGGGAACCTAACCGATTCC | 9732 |
| Qy | 2081 | HisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGluAla | 2100 |
| Db | 9733 | ATAATTACAGTACTACGGCACCAAT----- | 9759 |
| Qy | 2101 | AlaHisLeuProHisLeuArgPro-----LeuProGluSerGlnProSer | 2115 |
| Db | 9760 | -----CCTCATCTCGCGCCACCATATATGGCTATTTCGAGGAGACGCGATCAATT | 9810 |
| Qy | 2116 | SerSerPro-----LeuLeuGlnThrAlaProGlyValLys | 2127 |
| Db | 9811 | CTGCGCGCGGATCGCTGGAAGCAGAACCGTCTGTATGCAACAGAGAGCGCAGGAAG | 9870 |
| Qy | 2128 | GlyHis-----GlnArgValValThrLeuAlaGlnHisIleSerGluValIle | 2143 |
| Db | 9871 | CACCAATCCACAGCAGCAGCAGCAGCAACACAGCAGCAACATCACGCCCCAGCAGCAG | 9930 |
| Qy | 2144 | ThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPro--- | 2162 |
| Db | 9931 | CAGCAGCAGCAGCAGCAGCACCATGCCAGCAACATCATCCGCAGATCCCGGCACCTGGC | 9990 |
| Qy | 2163 | LeuTySerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp | 2182 |
| Db | 9991 | TCGTGCTCGGCTCCAGGTGGTCT----- | 10014 |
| Qy | 2183 | LeuTyLeuProProProAspHisGlyAlaProAlaArgLysProHisSerGluGly | 2202 |
| Db | 10015 | -----GGTCAGGGCGGT---GGATCGGGCGGCCCGGTAGCGGTGCT | 10053 |
| Qy | 2203 | Gly-----LysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAsp | 2220 |
| Db | 10054 | GGTGGCGCGCGCAGACCAACCT-----GGCGAGGAT | 10089 |
| Qy | 2221 | GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla | 2240 |
| Db | 10090 | GGCGGCAACATAATCCGAATGCCACG-----GGG | 10119 |
| Qy | 2241 | ValTyProLeuLeuTyArgAsp----- | 2248 |
| Db | 10120 | GTCACTCCCGGAAAAATTTAATCATGAGATGATGCTACATCATGTGATGGAAACGACGGGA | 10179 |
| Qy | 2249 | -----GlyGluGlnThrGluProSerArgMet----- | 2257 |
| Db | 10180 | CGCGGTGTGAAGCTGCTTCTTCTTACCCACCGCTGCTGTTCTGCTCCCGCAGCAGAGA | 10239 |
| Qy | 2258 | -----GlySerLysSerProGly-----AsnThrSerGlnPro | 2268 |
| Db | 10240 | GGAAAGCCCGCAGCGGTGGTGGAGCGCTGGAGCGCGGAGGTCTCGGAGGCGGT | 10299 |
| Qy | 2269 | ProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGln | 2288 |
| Db | 10300 | GCTACCACCATTTGAAAAATATGTAAGACCCGCAATTGCTGAAGTGTGCGCATATATT | 10359 |
| Qy | 2289 | GluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyArgAsnIleSerGln | 2308 |
| Db | 10360 | GGTTATGGAAGAATCGGACTGTGAGGTGCGAACAGAGGATGAAGTA----- | 10407 |
| Qy | 2309 | ProGlyThrLulilePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyArg | 2328 |
| Db | 10408 | -----ACCGCGCATATGTTGGCAGCAT | 10428 |
| Qy | 2329 | SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLys | 2348 |
| Db | 10429 | TCGCATCCGCCC-----GTCCATGTGCACATGTGGCGCACGACCCCATGTCGCCCATGCC | 10485 |
| Qy | 2349 | AlaLeuMetGlyLysTyArgGlnTrpGluGluSerProPro----- | 2362 |
| Db | 10486 | GCTGCTATGGAGTTGCGACAGCAAGCAAGAACCCACCGCCCGCCAGAGATCAGTGTGTCA | 10545 |
| Qy | 2363 | ---LeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMet | 2381 |
| Db | 10546 | CGTAAGACGCCCAACCAATACGAGGTGGTAGACGCCAGTGGCGACAGCGGTGGCACCA | 10605 |
| Qy | 2382 | ProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGly-----Gly | 2399 |
| Db | 10606 | CCTGGTGCCTTTCCCAAGCCCAAGCAGCGCATGCCCTGAGTGAATCGGTGCGTCCGT | 10665 |
| Qy | 2400 | GlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro | 2419 |
| Db | 10666 | GGTGGGTGTCTATTGGTGGGCGCGCGCTCTGGA----- | 10701 |
| Qy | 2420 | GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys | 2439 |
| Db | 10702 | GGAAATTCAGCGCGA-----CCAGGTGTGTCTCA----- | 10731 |
| Qy | 2440 | AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer | 2459 |
| Db | 10731 | ----- | 10731 |
| Qy | 2460 | ThrProPheProTyArgAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro | 2479 |
| Db | 10732 | -----GTCCGTGTCT----- | 10740 |
| Qy | 2480 | ProProGlyLeuProAlaGlySerGlyProLeuAla-----GlyPro | 2494 |
| Db | 10741 | -----GGTGTACCGCGCGCGGACCAAGAGCGGTGGCGCGCGGTGGCGGT | 10791 |
| Qy | 2495 | HisHis-----AlaTrpAsp | 2499 |
| Db | 10792 | CACATTCAGTTCATCCGAAGCATCTGCTGCGGTGCCGTGCGGTGGCGGCGGCC | 10851 |
| Qy | 2500 | GluGluProLysProLeuLeuCysSerGlnTyArgGlnThrLeuSerAspSerGlu | 2517 |
| Db | 10852 | AGTCAGTCCAAGCGCTGCTGTTCTCCAAGTACATGATGCTACGACGACGAAGAT | 10905 |

5225 TGCCGCTATTGGCGGATGGCAGTGGCAACTCGATGCTCAATGCAACAGAGATTCTCGCCC 5284
QY
782 -----ProProT 784
Db
5285 TCAGTGGCAAGACAAATTGGCCAGCTGCTTTGTGTGCAAGGCGGAGGCGTGTCCCGCA 5344
QY
784 hrProProArgArgThrSerArgAla-----ProLleGluProThrProA 799
Db
5345 CCCTGCACTAAAGAGGGCGTGGCCAGCAGTATGGCATTCGGACGAAACAATACCGG 5404
QY
799 laSer-----GluAlaThrGlyAlaProThrProProAlaP 812
Db
5405 CGGTGCGCCGCTGCAATAGCTGCAATGCAATCGGTGCGCAGCGCTATCCAAAT 5464
QY
812 roProSerProSerAlaProProVal-----valProL 824
Db
5465 GTCCGCTACCCATGCCCAATCCCAAGGATCGAGCCCGCTCTCGGAAACATTCCT 5524
QY
824 ysGluGluLysGluGluThrAlaAlaProProValGluGlu-----GlyG 841
Db
5525 CCGCTCTCTCGAA-----CTGGCGCCGAGGTGCGTATCGGTGATCGCGC 5572
QY
841 luGluGluLysProProAlaAla----- 848
Db
5573 AATTCAAATACCACCGCATGCAACACGCTGCTGCTCGGCTGCTAATGCGCATTCGTC 5632
QY
848 ----- 848
Db
5633 GCAAGCTCGATCCGCAGCTGAATCTCACGATGGCTCCAGTGGCGTGGCGGAGCGGCA 5692
QY
849 -----GluGluLeuAlaValAspThrGlyLysAlaGlu-----GluProValLys 864
Db
5693 GTGGCGCATGACGGATGTAGTACCTCTGCTGCTGATGACGAGCCGCGGCGAT 5752
QY
864 erGluCysThrGluGluAlaGluGlu-----ProAlaLysG 877
Db
5753 CCGATCCGCTTCGTTGA--GAGTCCGAGAAATTGTCAGCGGCAAGTCCCTTAGCATG 5811
QY
877 lyLysAspAlaGluAlaGluAlaThrAlaGluGlyAlaLeuLysAla--GluLysLys 896
Db
5812 GTCAACAGCAACAGCAACAGCAGCAACAGCAGCAGCAGCAACAGCAGCAGCAGCAG 5871
QY
897 Glu----- 897
Db
5872 CAACAGCAACAGCAACTATCGCAGCCACGCCACCTCCGCCAGTCCGCGCAGCAGCAAG 5931
QY
897 ----- 897
Db
5932 GGTTCGTCTGGCAGGCTGTGATCAGGGTACACCACTGATTATCACCAACCCGAATG 5991
QY
898 -----GlyLysSerGlyArgAlaThrAla----- 906
Db
5992 AGCAGCAAGTCGGGATCGGGCGGTGCGCAACAGCGGGCGACAAATGAACGCTACTGCCA 6051
QY
907 -----LysSerSerGlyAlaProGlnAspSerAspSerAla 919
Db
6052 CCGCCGCTGCCAGGCAACCCAGAACAGCAACAGCAGGAGGAATACGATTCTCGGCC 6111
QY
920 ThrCysSerAlaAspGluValAspGlu----- 928
Db
6112 ACGGAACCGCGGATGAGGAGAACGAAATCTCGCGGCCAATCGTCAGAGTCCCAAGGTA 6171
QY
929 -----AlaGluGlyClyAspLysAsnArgLeuLeuSerProArg 941
Db
6172 CTATTCCATGGCCATGGACATGGCATGGTGGTTCATGCCAATATGTGGTGTCTCGAC 6231
QY
942 Pro-----SerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGln 958
Db
6232 CCGCGGTGGCCATATGGGACGGGTGGTGTTCACACGAGTGGAGCTGTGGGCAA 6291
QY
959 Lys-----ProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaLeuProPro 975
Db
6292 CAGGTCAACGAGCCGATTAGTATG-----CGACCGGCGGCGGTCAACAAT 6336

QY
976 IleGln-----ValThrLysValHisGluProProArgGluAspAlaAlaProThrLys 993
Db
6337 GTTCAGATTGGTATTCTCGTAATTGAG-----CGTTCGCTGAAG 6378
QY
994 ProAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGln 1013
Db
6379 CACAAGGGTCCGCAACCGAAGGCGGACAGGTGTAGCAGCAGGG--TCAGGGCAAGGACA 6437
QY
1014 GlnProGlySerProArgGlyLysSerArgSer---ProAlaProProAlaAspLys 1032
Db
6438 GGGCCAGGTGTAGGTTCAGACTCCAGGTCAAGCCAGTCCCATCCACAGCAACAGCA 6497
QY
1033 GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGly 1052
Db
6498 GCAGCAGCAGCAGCATCTGCCAATACTGGAACCGCAGGAGCTTACCATGTTAAGGA 6557
QY
1053 LeuProPheProValPro--ProArgGluValLleLysAlaSerProHis----- 1068
Db
6558 ATACCG-----CCAGGATCCGGGTATACTGAAGCAGCAGCAGCAGCAGCAGCAGC 6608
QY
1069 -----AlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLe 1086
Db
6609 TGGCGAGCAGCCGCCACCTCGGCTGCCGTAGTTTCCGCGATGGC----- 6654
QY
1086 uGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPr 1106
Db
6655 -----ACCTCGTACAGAAGCTAACACCCGCGCCAGCTGCAGTGGCACCCAGCC 6704
QY
1106 oProLeuLleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaLleSe 1126
Db
6705 ACCG-----GCTCATCCA---CTGCACCCACCCAGCAGCAGTGGATGTCGGG 6746
QY
1126 rGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValG 1146
Db
6747 CAGCAACACCGAACCAGCGATAGCTTACGCCACCTTATCGGTGCTCAATTCGCACATGG 6806
QY
1146 yProValThrMetGlyLeuLeuProLeuProMetAspProLysLysLeuAlaProPheSerG 1166
Db
6807 CATGTGGGAATTGGTCATCCGCGCCGATGGCCACGCAAGTTCCCGCGAGGAATCGG 6866
QY
1166 yValLysGlnGluLleuSerPro-----ArgGlyGlnAlaGlyProProGluSerLe 1184
Db
6867 TGTGACAAAGCGGCACCATTAACGCCGTTGTCAAGAGCTCTTAGTGGATCATCGAAGAGCG 6926
QY
1184 uGlyValProThrAlaGlnGluAlaSerValLeu---ArgGlyThrAlaLeuGlySerVa 1203
Db
6927 CGCGCGCAGTGGTCCAGTCATCTGCAGCCACACCGCCGAGACCAATAATCTACATGT 6986
QY
1203 lProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerSerAspSerAla 1223
Db
6987 GCCAGTGGCGCATCCGCAACGTGGTATACCA-----CCGTGCGCAGCAGCTCGGT 7037
QY
1223 e-----ThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysG 1241
Db
7038 ACATCGGCCCATCCGTCGCATACAGCATCCGCCCATCCGCGCCATCCGAGCATTCGTCGATGG 7097
QY
1241 yThrIleThrArgLle---IleGlyGluAspSerProSerArgLeuAspArgGlyArgG 1260
Db
7098 TCAGCATACCACTGCAAGTGGCGAACCAGGAGCGCAACATTGGACTTGGATGACATCA 7157
QY
1260 uAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTy 1280
Db
7158 GAAGCGCGCGCGATGGCCATTCCGCGCACACGGGA----- 7194
QY
1280 rGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSer----- 1298
Db
7195 -----GCCGAGGATCTAGCTCTGCTGCTGATCGGGTGGACCGCTCTCATCGGATCG 7250
QY
1299 -----GlyProHisGluThrAlaAlaProLys-----ArgThrTyrAspMe 1313
Db
7251 ACATCATGGCCACCGCGGACCATGTGATGAGACACATGTTGACGCTCTGGTGGCAT 7310

Db 422 CAGCGGAGCTGCAGAGCGCATGCAGACGAGGTGGCGGCGAGTGGCTGTGCC 481
 Qy 363 MetSerAlaAlaArgSerGluHisGluValSerGluLeuLeuLeuLeuLeuLeu 382
 Db 482 ATGTGGCGCGCGCAGGAGCAGGAGTGTGAGAGATCATGATGGCTCTCAGAGCAG 541
 Qy 383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValLeuProMetLeuTyrAspAla 402
 Db 542 GAGAACCTGGAGAGCAGATGCGCCAGCTGGCGGTGATCCGCCCATGTGTACGACGCT 601
 Qy 403 AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAsp 418
 Db 602 GACAGCGAGCGATCAAGTTCATCATCATGATGAGCGGCTTATGGCCGAC 649

RESULT 13

ABL03131
 ID ABL03131 standard; cDNA; 10910 BP.

XX ABL03131;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 3875.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.
 XX P-PSDB; ABB59028.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions.

XX Claim 1; SEQ ID NO 3875; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABU16176-ABL3051), expressed DNA
 XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 XX ABB72072). The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 10910 BP; 2875 A; 3179 C; 3157 G; 1699 T; 0 U; 0 Other;

Alignment Scores:

| | | | |
|------------------------|----------|---------------|-------|
| Pred. No.: | 1.07e-24 | Length: | 10910 |
| Score: | 1083.00 | Matches: | 724 |
| Percent Similarity: | 33.03% | Conservative: | 360 |
| Best Local Similarity: | 22.06% | Mismatches: | 1076 |
| Query Match: | 8.20% | Indels: | 1128 |
| DB: | 4 | Gaps: | 137 |

US-09-522-753-5 (1-2517) x ABL03131 (1-10910)

| | | | |
|----|------|--|------|
| Qy | 20 | ProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu | 39 |
| Db | 2092 | CCACCACAC-----CCGACGAG-----CAACACACGACGACGACGACG | 2130 |
| Qy | 40 | LeuGluTyrGlnHis-----HisSerArgAspTyrAlaSerHisLeuSerProGlySer | 57 |
| Db | 2131 | CAGCAGACGACACATCAGCGCGCACGACGACCAACAT---CAGCGATCTCTCAGCAGCA | 2187 |
| Qy | 58 | IleIleGlnProGluArgArgProSer----- | 67 |
| Db | 2188 | GTGCAACACGACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG | 2247 |
| Qy | 68 | -----LeuLeuSerGluPheGlnProGlyAsn | 76 |
| Db | 2248 | CGACAAAGTGTGCACACGCCAATGGCATGTCAGCTCAAGTGGAGACGCGTCCACGACG | 2307 |
| Qy | 77 | GluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly | 96 |
| Db | 2308 | CAACAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG | 2361 |
| Qy | 97 | LysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspPro | 116 |
| Db | 2362 | CGATCTCAG-----CCAGTGTGAGCAGCATGTCCACCGTG | 2397 |
| Qy | 117 | LeuLeuArgPro----- | 120 |
| Db | 2398 | GTGAGCCAAACAGTGGCAGCGTTACCGTTACCAGCGCGGATTATCGCGGACGACACG | 2457 |
| Qy | 121 | -----SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp | 138 |
| Db | 2458 | GGCAGCAGTGGCAATGTTCCCGCGGATTGGGTGCGGAAATACGGGATCCCGCAGCACC | 2517 |
| Qy | 139 | ArgSerLeuThrGlyLysLeuGluProValSerProProSerProHisThrAspPro | 158 |
| Db | 2518 | GAGCCCTATCCTCCTCAGTGGAGGCCATTTGCCGACATGCCC-----AGCGATAGT | 2571 |
| Qy | 159 | GluLeuGluLeuValProArgLeuSerLysGluLeuLeuIleGlnAsnMetAspArg | 178 |
| Db | 2572 | TCGATTGAGGAGCGCGGAAGGACCGACGCGCAAGAGGATCTTCTCATCAAAATCCAAAG | 2631 |
| Qy | 179 | ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGln | 198 |
| Db | 2632 | GTTCACAAATGAGATCAAAATCCGCTGAGACGACACATGGAACGTTGCGCAAAAGGAGAA | 2691 |
| Qy | 199 | GlnLeuGluGluAla-----AlaLysProProGluProGluLysProValSer--- | 215 |
| Db | 2692 | TCCTTCATGAGGAGCGCGCCCTGCGCCAAAGACGACGAGGCGCCCAAGAGTTGAACG | 2751 |
| Qy | 216 | -----ProProIleGluSerLysHisArgSer-----LeuValGln | 228 |
| Db | 2752 | AACAACAATGATCAGGAACCGATGTTGGAATCTATCGTGGCGGACGCAAAATGCTAGCGG | 2811 |
| Qy | 229 | IleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeu | 248 |
| Db | 2812 | AAATCTATGCGGCAATCGGAAGACGCGCCCAAGCGCAACATTCATGTCGAGAATGCA | 2871 |
| Qy | 249 | GlyProGlnValGlu----- | 258 |
| Db | 2872 | GCGGCGGATGAATCGTCTCGGGTTCGGTTGCGCGGACGACCATGTTGCTTGTATTAAC | 2931 |
| Qy | 259 | GlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAla---MetArg | 277 |
| Db | 2932 | CAGCCATCGATGTCGAGGCACCTTGCATGCTAATAGCGGACGATCAGATCAAAATTCGT | 2991 |
| Qy | 278 | LysLysIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLys | 297 |
| Db | 2992 | GCGCGCTACTTTTGACATACAAATTAAGCGCCGACCGTGGCGGCAACATCAGGGA | 3051 |
| Qy | 298 | PheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGlu | 317 |
| Db | 3052 | CTGCTGGAGAAGTACACCAAGGATCAGCGCGGATTGGCAGCGGACGATGCGGATGGAG | 3111 |

Db 633 AATGACCTGAATACAAATATACACCGCTGGAGCAGATCTTCAATATGCCGCCATC 574
 QY 2320 ThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsn 2339
 Db 573 ACCGGAC-----AGAGCCAGCGGTGCAGGAACATGCCAGCACCAAC 530
 QY 2340 MetGlyLeuGluAlaIleleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu 2359
 Db 529 ATGGGCTGGAGGCCATAATAGAAAGGCACCTCATGGTAAATATACCAAGTGGGAAGAG 470
 QY 2360 SerProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla 2379
 Db 469 TCCCCCGCGCTCAGCCCAATGCTTTAACCTCTGAATGCCAGTCCAGCCGCGCT 410
 QY 2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2399
 Db 409 GCTATGCCCATACCGCTGTCAGCGACGAGGAGTGACACACACTCACTCGCCAGGTGGC 350
 QY 2400 GlyGlyAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro 2419
 Db 349 GCGCGGAAGCCCAAGGTCTCTGGCAGACCCAGCAGCGAAAGCCAGCCCGCCAGCC--- 293
 QY 2420 GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys 2439
 Db 292 -----CGCTCCGGGTGGAGCCCGCTCTCCCGCAGCAAGGGCTCGGAG----- 248
 QY 2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
 Db 247 -----CCCGGCCCTAGTG-----CCTCTGCTCTGCGCAC 215
 QY 2460 ThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
 Db 214 GCCACCATCGCCGCCACCCCT-----GCGAAGAACCTCGCACCTCACCAAGCGCCGCG 161
 QY 2480 ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisAlaTrpAsp 2499
 Db 160 GACCCCGCGCGCCACTCGCTCG-----GCTCGGACCCGCGCACCGGGAAG 113
 QY 2500 GluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSer 2514
 Db 112 ACTCAAGTAAACCTTTTCCATCCAGGAACCTGGAACCTCGTTCT 68

RESULT 12
 ACA57401
 ID ACA57401 standard; cdna; 650 BP.
 XX
 AC ACA57401;
 XX
 DT '10-JUN-2003 (first entry)
 XX
 DE Human adipocyte Selected Interacting domain, SID, cdna #488.
 XX
 KW Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
 KW anorectic; antidiabetic; protein-protein interaction; diabetes;
 KW yeast 2-hybrid assay; metabolic disorder; obesity.
 XX
 OS Homo sapiens.
 XX
 PN W0200286122-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 14-MAR-2002; 2002WO-EP003768.
 XX
 PR 14-MAR-2001; 2001US-0275734P.
 XX
 PA (HYBR-) HYBRIGENICS.
 XX
 PI Legrain P, Daviet L;
 XX
 DR WPI; 2003-103412/09.
 DR P-PSDB; ABU70857.
 XX

PT New complex between two interacting proteins in adipocyte cells, useful
 PT for identifying selected interacting domains that modulate protein
 PT interactions, or for preventing or treating metabolic disorders such as
 PT obesity or diabetes.
 XX

PS Claim 7; Page 271; 382pp; English.

XX The invention relates to a complex between two interacting proteins in
 CC adipocyte cells, given in the specification. The proteins are identified
 CC by selecting a bait protein from a known adipocyte marker and then
 CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
 CC members of an adipocyte cDNA library. The proteins are designated SID
 CC (RTM) (selected interacting domains) proteins. Also included are a
 CC polynucleotide encoding a polypeptide in the adipocyte cells, a
 CC recombinant host cell expressing at least one of the interacting
 CC polypeptides of the complex, selecting a modulating compound in adipocyte
 CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
 CC sequences given in the specification (including its fragment or variant),
 CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
 CC given in the specification (including its fragment or variant), a vector
 CC comprising the SID (RTM) polynucleotide, a recombinant host cell
 CC comprising the vector, a protein chip comprising the polypeptides and a
 CC record comprising all or part of the data, listed in the specification.
 CC The complex, polypeptides, polynucleotides and compounds are useful for
 CC preventing or treating metabolic disorders such as obesity or diabetes.
 CC The polynucleotides are useful as probes or primers. The complex is
 CC particularly useful for identifying selected interacting domains (SID
 CC (RTM)) for screening drugs that modulate the protein interaction, thus
 CC exhibiting the therapeutic effect. The present sequence encodes a SID
 CC (prey) protein of the invention

XX SQ Sequence 650 BP; 170 A; 186 C; 205 G; 89 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,48e-26 Length: 650
 Score: 113.00 Matches: 216
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.42% Indels: 0
 DB: 7 Gaps: 0

US-09-522-753-5 (1-2517) x ACA57401 (1-650)

QY 203 GluAlaAlaLysProGluProGluLysProValSerProProProfileGluSerLys 222
 Db 2 GAGGCTGCCAAGCGCGCCGAGCTGAGAAGCGCTGCACCGCGCCCATCGAGTCCAAG 61
 QY 223 HisArgSerLeuValGlnIleTyrAspGluAsnArgLysLysAlaGluAlaHis 242
 Db 62 CACCGCAGCTGTGTGCAGATCATCTACGACGAGAACCGGAAGGCTGAAGCTGCACAT 121
 QY 243 ArgGluLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAsp 262
 Db 122 CGGATTTCTGGAGGCTTGGGGCCCGAGGTGGAGTCCCGCTGTACAAACGACCTCCGAC 181
 QY 263 ThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuLeu 282
 Db 182 ACCCGCAGTATCATGAGAACATCAAAATAAACCCAGCGCATGCGGAAGACTAATCTTG 241
 QY 283 TyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyr 302
 Db 242 TACTTCAAGAGGAGGAATCACGCTCGGAACCAATGGAAGCAGAAAGTTCTGCCAGCGCTAT 301
 QY 303 AspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArg 322
 Db 302 GACCACTCATGGAGGCTTGGNAAAAAAGGTGGAGCGCATCGAAACACCCGCGCGG 361
 QY 323 ArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLys 342
 Db 362 CGGGCCCAAGGAGCAAGGTGCGGAGTACTACGAAAGCAGTTCCCTGAGATCCGCAAG 421
 QY 343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer 362

CC cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung
CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,
CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary
CC thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,
CC Addison's disease, thyroditis, Crohn's disease, Graves' disease,
CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.
CC viral, bacterial, fungal or parasitic infection), developmental disorders
CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.
CC thymomas, hypopituitarism, hypogonadism, gigantism, goiter), metabolic
CC disorders (e.g. hypercholesterolemia, hypoglycaemia, diabetes,
CC hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic
CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,
CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,
CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),
CC transport disorders (e.g. akinesia or multidrug resistance), or
CC connective tissue disorders (e.g. Paget's disease or rickets). This
CC polynucleotide sequence represents one of the human dithp DNA sequences
CC of the invention.

XX
SQ Sequence 2336 BP; 367 A; 634 C; 894 G; 441 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,36e-49 Length: 2336
Score: 171.50 Matches: 427
Percent Similarity: 57.40% Conservative: 19
Best Local Similarity: 54.95% Mismatches: 99
Query Match: 14.16% Indels: 233
DB: 9 Gaps: 14

US-09-522-753-5 (1-2517) x ADE31306 (1-2336)

Qy 1935 ProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAsp 1954
Db 2286 CCACTTCTCACTCGG-----CCGCACACACTCTCTCCCGACAGACCCATGTCGCC 2236

Qy 1955 ThrGlyHisAlaPheLeuAlaLeuProAlaArgSerGlyLeuGluProAlaSerSer 1974
Db 2235 CCATCTCAGCTGGCTGGCGCGGCTCCACCACT-----GAACCACTCTGTCTC 2185

Qy 1975 ProSer-----LysGly-SerGluProArgProGluValProValSerGlyHisAl 1992
Db 2184 CCTCTTGGCTAAATCTCTCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2125

Qy 1992 aThrIleAlaArgThrProAlaLysAsnLeuAla----- 2003
Db 2124 ATCTCTCTCCACTCTCTCTTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2065

Qy 2004 -----ProHisAlaSerPro-AspProAlaProp 2015
Db 2064 TGTCTTTGGGAGCCCAACCTACCTAGCTTGGTTCCTTGTGACCCCGCGGGTCC 2005

Qy 2015 roAlaSer-----AlaSerAspProHisA-gGluLysThrGlnSerLysProPhes 2032
Db 2004 CAGCAGCTGGAGGAGCCCTCGCCCTCGGCTCCGAAACCTCGGCGCGGTGCTGA 1945

Qy 2032 erIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
Db 1944 CTCTGCACCCCGCGCT-----GCCCTAGTGTACCCAGGAGGAGCTACAGCCCCGAAG 1891

Qy 2052 lyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeuProL 2072
Db 1890 GGTGGAGCCCGTCTAGCTGTGAGTCTACCCAGTCTGACCCACGACGAGGGGTCCCA 1831

Qy 2072 ysHisLeuGluLeuLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG 2092
Db 1830 AGCACCCTGGAGAGCTCCACAAGACCCCTCGAGGGGAGCTGGCGGCCAAGCAGCCAG 1771

Qy 2092 lyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLysGln 2112
Db 1770 GCCCGGTGAAGCTTGGCGGGAGGCGCCCACTCTCCACACTCGCGCGCTGCTGTAGA 1711

Qy 2112 erGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgv 2132

Db 1710 GCCAGCCCTGTCTCAGCGCGCTGTCTCAGACCGCGCCAGGGGTCAAGGTCAACGCGGG 1651
Qy 2132 alValThrLeuAlaGlnHisIleSer----- 2140
Db 1650 TGGTCACTTGGCGCCAGGACATCACTGTAACTACCGCTTCTGTCTGTCTGTCTGTCACT 1591
Qy 2140 ----- 2140

Db 1590 TTGCACCTGGGGGACCAGGCTTGGAGAGGGATGGGAACCCACAGACCTTCTGTCTCT 1531
Qy 2140 ----- 2140

Db 1530 GCGGGGTGCTGGGGATCCAGGCGATGCGCTGGGGGATCCAGGGCGTGGGTGAGGG 1471
Qy 2140 ----- 2140

Db 1470 TGAGATCCAAAGCCCCGAGACCGGACCATCATCCGCCCTTAATCCATGGGAGAGCC 1411
Qy 2140 ----- 2140

Db 1410 TGTGATGCCAGCGGATGGCATCTTCACGGGCAATGAGGCTTCTTGTGGCCCGAGTTTC 1351
Qy 2140 ----- 2140

Db 1350 TCAGTGTCTATGGCTGGTCTCATCAGCCATCTGCCAACTACCAGTCTGGGACCGGTGACC 1291
Qy 2140 ----- 2140

Db 1290 ACAGCCCCACTCCATGTCACATGGGACACGAGGCCCCAGAGGTGGCGGCGAGTCCAC 1231
Qy 2140 ----- 2140

Db 1230 AGTCACCCAGAGCTGGCCCCCAGGATCTGCCCCGAGTCTCGCTCTAGCCCCCTCCC 1171
Qy 2140 ----- 2140

Db 1170 CACCCCGCAGAGGTTCTGTCTCAGGAGAGTGTCTGCTGACTCTGGGCCCCCCCCACTTGCCTG 1111
Qy 2141 ---GluValIleThrGlnAspTyrThrArgHisProGlnGlnLeuSerAlaProLeu 2159

Db 1110 CAGAGGTATATCACAGGACTACACCCGGCACCAACAGAGCTCGGTCTTGGTGGTGTGAG 1051
Qy 2160 ProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgPro 2179

Db 1050 CGCCCCCTCTACTC---CTTCTGGGCGCAGCTGCCCCGCTCTGACCTCCGCGGCCA 994

Qy 2180 ProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHis 2199
Db 993 CCCAGTGACCTTACTCTCCCGCCCGGACCATGGTGCCCGCCCGCTGCTCCCCCACC 934

Qy 2200 SerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGlu 2219
Db 933 AGCGAGGGGCGAAGAGTCTCCAGAGCCAAACAGAGCTCGGTCTTGGTGGTGTGAG 874

Qy 2220 AspGlyIleGluProValSerProGluGlyMetThrGluProGlyHisSerArgSer 2239
Db 873 GACGGTATTGAACCTGTGTCCCCACCGAGGCGATGACGAGGCGAGGCGCATCCCGGAGT 814

Qy 2240 AlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySer 2259
Db 813 GCTGTGTACCGCTGTGTACCGGATGGGGAACAGAGGAGCCACAGAGTGGGTCTCC 754

Qy 2260 LysSerProGlyAenThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsn 2279
Db 753 AGTCTCCAGCAACACAGCCAGCCAGCCAGCTTCTTCAGAGCTGACCGAGAGCAAC 694

Qy 2280 SerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAenThrHisAsnArg 2299
Db 693 TCCGCCATGTGTCAAGTCAAGAACAGAGATCAACAGAGCTGAACACCCACACCGG 634

Qy 2300 AsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIle 2319

Db 6012 CTCTCTACTTCTACATTCACAGAACTCACCTTCTGCTTTGGTATCTACACCTGTG----- 6065
Qy 2176 LeuArgProProSerAspLeuTyLeuProProAsp-----His 2190
Db 6066 ---AGGACAAACATCAACCGTTACAGCCAGAAATCCAGAGCTCAGTCTGTCATCAT 6122
Qy 2191 GlyAlaPro---AlaArgGlySerProHis-----SerGluGlyGlyLys 2204
Db 6123 CAAAGACCAAGGTTCAAGGGTCTCTCCAGAAATCTTGTGACAAATCCAGGGGAAGTAGG 6182
Qy 2205 ArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyLleGluPro 2224
Db 6183 CTGGGAAATCCCGAGAGGAGTCCAGTCT-----TCTTCGGAGCCCTACGAGCCC 6233
Qy 2225 ValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyProLeu 2244
Db 6234 ATCTCCCCACCCAG-----GTTCCGGTGTGTCATCAGAAACAGCAGCTTGTGCTC 6287
Qy 2245 LeuTyArgAspGlyGluInThrGluProSerArgMetGlySerLysSerProGlyAsn 2264
Db 6288 TTGTCTCAGAGGGCGCAGAGCTCGCAGCAGAGGAATGATGCCGCTCACCAGGGAGT 6347
Qy 2265 ThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLys 2284
Db 6348 ATAAGCTACTTGCCTTCTTCAACCAAGCTT---GAAATATACATCACCCATGGTTAA 6404
Qy 2285 SerLysLysGlnGluLeuAsnLysLeuAsnThrHisAsnArgAsnGluProGluTy 2304
Db 6405 TCAAGAGACGAGAGATTTTCTGAGTTGAACTCTCTGCTGGAGGTGACTCTGATATG 6464
Qy 2305 AsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeu 2324
Db 6465 GCAGCTGCTCAGCAGCACTGAGATCTTAACTGCCAGCAGTACTACGTCAAGCTCA 6524
Qy 2325 MetThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAla 2344
Db 6525 GTTAGCTCTAGAGCCATCTTTTGTGTCATCCTGCCAGT---AATCTGGGCTGGAAGAC 6581
Qy 2345 IleIleArgLysAlaLeuMetGlyLysTyArgGlnTrpGluGlu----- 2359
Db 6582 ATTATCAGGAGGCTCTCAGGAGCTTTGATGACAAAGTTGAGGATCATGAGTTGTC 6641
Qy 2360 ---SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro 2378
Db 6642 ATGCCAGCCTATGGAGTA-----GTGCTGGTACTGCCAACACCTCAGTT--- 6689
Qy 2379 AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGly 2398
Db 6690 -----GTGACCACTGTGAGACACGAGAGAGAGGAGGAGGCCATCACCTCAT 6737
Qy 2399 GlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerPro 2417
Db 6738 TCAGGAGTTTGCAACCAAGCTGATCAGCAAGTCAACAGCAGAGGAATCTAGTCTCT 6797
Qy 2418 AlaPro-----GlyLeuAlaSerGlyAspArgProProSerValSerValHisSer 2435
Db 6798 ATACCTGGGCAAGCTACTTAGGAACGAAACGGCCCTCTTTCAGTCTCTCTGTACATCA 6857
Qy 2436 GluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSer 2455
Db 6858 GAAGGGGATTTACCATGGCAGAGCCCA-----GGGTGGGCCCTGGGAAGACAGGCCCTCT 6911
Qy 2456 SerAlaGlySerThrProPheProTyArgProLeuIleMetArgLeuGlnAlaGlyVal 2475
Db 6912 TCAACAGGCTCACTGAGTTTCTTATACCTCTGACTATGGGATG-----CTC 6962
Qy 2476 MetAlaSerProProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyPro 2494
Db 6963 AGCAGTACTCCACCAACCGATTCGATGTGCTCCCTCTCGCGTGAACCAAGCAGCTCCT 7022
Qy 2495 His-----HisAlaTrpAspGluProLysPheProLeuLeuCysSerGlnTyGlu 2511
Db 7023 CACCAACAGAACAGGATCTGGGAGCGAGAGCCCTGCCCACTGCTCTCAGCACAGTAGCAG 7082

Qy 2512 ThrLeuSerAspSerGlu 2517
Db 7083 ACCCTGCGATGATGAT 7100

RESULT 11

ID ADE31306/c

XX ADE31306 standard; DNA; 2336 BP.

AC ADE31306;

XX 29-JAN-2004 (first entry)

XX Human diagnostic and therapeutic polynucleotide (dithp), SEQ ID No 61.

XX diagnostic and therapeutic polynucleotide; dithp; antiarteriosclerotic;
KW antiinflammatory; cerebroprotective; antilipemic; antidiabetic;
KW immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;
KW osteopathic; antiarthritic; antirheumatic; cytostatic; hepatotropic;
KW virucide; haemostatic; anti-HIV; antithyroid; thyromimetic;
KW dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;
KW thrombolytic; anticoagulant; anorectic; vasotropic; antiulcer;
KW gene therapy; protein replacement therapy; human; gene; ds.

XX Homo sapiens.

XX WO2003062376-A2.

XX 31-JUL-2003.

XX 13-JAN-2003; 2003WO-US001096.

XX 16-JAN-2002; 2002US-0349384P.

XX 17-JAN-2002; 2002US-0349413P.

XX 17-JAN-2002; 2002US-0349946P.

XX (INCY-) INCYTE GENOMICS INC.

XX Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL;

PI Yu JY, Tusson O, Yap PE, Ameshey SR, Dam TC, Liu TF, Gerstin EH;

PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RY, Urashka ME;

PI Kristnam SR, Kolluru V, Panesar IS;

XX WPI; 2003-636732/60.

XX P-PSDB; ADE31117.

XX New human diagnostic and therapeutic polynucleotides and polypeptides,

PT useful for diagnosing, treating or preventing e.g. leukemia, brain

PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke

PT or Alzheimer's.

XX Claim 1; SEQ ID NO 61; 634pp; English.

XX The invention relates to a novel isolated human diagnostic and

CC therapeutic polynucleotide (designated dithp). The novel dithp

CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798

CC base pairs fully defined in the specification; a polynucleotide

CC comprising a naturally occurring polynucleotide sequence at least 90%

CC identical to the dithp polynucleotide; a polynucleotide complementary to

CC the dithp polynucleotide or its polynucleotide which is at least 90%

CC identical; or an RNA equivalent of any of the polynucleotides mentioned

CC above. The dithp polynucleotides have the following activities:

CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipemic,

CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,

CC tranquilizer, osteopathic, antiarthritic, antirheumatic, cytostatic,

CC hepatotropic, virucide, haemostatic, anti-HIV, antithyroid, thyromimetic,

CC dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,

CC thrombolytic, anticoagulant, anorectic, vasotropic, and antiulcer. The

CC novel DITHP polynucleotides polypeptide can be used in gene therapy and

CC protein replacement therapy. The dithp polynucleotides or DITHP

CC polypeptides are useful for diagnosing, preventing or treating diseases

CC associated with the expression of human molecules. In particular, these

CC diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain

1506 QY -----GluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGly 1520
1404 DB GATGTGAAGACGCGGACCGTGTCCCGGACACAGTCAGTGTAGCTCTGCGCCC 4163
1521 QY SerIleAlaArgGlyAlaProValIleProGluLeuGlyLysProArgGlnSerPro 1540
1464 DB TCGGTCTTAGTCCACA-----CTGCATGAAGCTCCCAAGACCAACTGAGCCCT 4214
1541 QY LeuThrTyrGluAspHisGlyAla-----ProPheAlaGlyHisLeuPro 1555
1421 DB GGGATTTATGATGACACAGCAGTGCAGCGGAGGCCCTGTGAGTTATCAAAACACCATGTCC 4274
1556 QY ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSer 1575
1475 DB AGAGGCTCACCCATGATGAACAGA-----ACTTCTGATGTTTCT 4313
1576 QY SerSerLysAla-----SerGlnAspArgLysLeuThrSerThrProArgGlu----- 1591
1431 DB TCTAACAGTCTACCAATCATGAAGGAATCGACACTGACCCCTACCCAGAGGGAAGT 4373
1592 QY ---IleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPro 1610
1474 DB ATCCAGCGAAGTCTCCAGTGCCTCGGGTGGACCTCTCGTGAGCCAC-----AGTCCG 4427
1611 QY TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeu 1630
1428 DB TTTGATCCCCATCACAGGACGACCTGCGAGCGAGTTATTCGAGCCACTGCCACG 4487
1631 QY AlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyr 1650
1448 DB CACTTGGATCCA---GCCATGCCCTTTTACAGGGCTTTGGATCTCGCGCTGTCTTACCTG 4544
1651 QY LeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIle 1670
1454 DB TTTCAGAGACAGCTTTTCCCAACTCCAGGTTTACCAAGTCAGTATCAGCTTTAC----- 4598
1671 QY ArgGlyTyrProAspThrAlaAlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyr 1689
1499 DB -----GCAATGGAGAACACAGACAGACACATCTTAATGATATAC 4637
1690 QY IleThrSerGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMet 1709
1438 DB ATTACCTCACAAAGATGCAAGTGAATCTG-----CGTCCAGATGTG 4679
1710 QY LeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArg 1729
1480 DB GCCAGAGACTCTCCCAAGAGAGCAGCCTGGTCTCCCATCCAGCA---ACGAGA 4736
1730 QY GlyIleLeuAspLeuSerGlnValProHisLeuProValLeuValProThrProGly 1749
1473 DB GGAATCATTTGACCTGACCAATATGCTTCCA---ACAATTTTGTGCTCATCCAGGGGA 4793
1750 QY ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer 1769
1494 DB ACAAAGCACTCTCCCATGGACAGATCACTTATATCTTCTGTGTACACAGATTACTTTCCCT 4853
1770 QY SerArg---HisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPro 1788
1484 DB CCCAGGCGTACAATCTGCTTCCATGTCTCCAGGACACCCACACACCTT----- 4904
1789 QY ThrThrThrSerSerSerGluArgGluArgAspArgArgGluArgAspArg 1808
1495 DB GCAGCTGTGCAAGTGTGAGAGGGAACGGGAACGGAGCGGAGAGGAGCGGAGCGG 4964
1809 QY GluArgGlyLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArg 1828
1465 DB GAAACGG-----ATTGCTGCAGCTTCTCCGACCTC-----TACCTGCGG 5003
1829 QY ProGlyThrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlySerSer 1848
1504 DB CAGGCTCTCAGAACAG-----CCT 5021
1849 QY SerArgProAlaSerHisHisAlaHisGlnHisSerProIleSerProArgThrGln 1868

5022 DB GCGCGACCTGGCAGTCATGGATATGTTGCTCCCTTCCCTCT---TCAGTAAGAACTCAG 5078
1869 QY AspAla---LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIle 1887
5079 DB GAGACCATGTTGCAACAGAGCCAGCTGTTTCCAAAGAACCAATGGAACCACTGTAATC 5138
1888 QY ThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerProVal 1907
5139 DB ACACCTTTGGATCCAACCTGCTCAGTACGAATCATGCCACTGCTGCTGGGGCCCTTCA 5198
1908 QY -----ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGly 1923
5199 DB ATAAGCCAAAGCGCTGCGAGCTCCCGTTACAACTGCTGCGGATGCC---CTGGCTGCT 5255
1924 QY ThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLys----- 1940
5256 DB CTGTGTGATGCTGCGAGCTTCTGACCCAGAGTGGATGTCTCCAAACAAAGAGAGTAAG 5315
1941 QY ---GluAlaProArgValAlaAlaArgProGluArgProArgAlaAsp----- 1954
5316 DB CATGAAGCTGCCAGGTTAGAGAAATTTGAGAGAGCAGCTCAGCAGCAGTTAGTGAACAG 5375
1955 QY -----Thr 1955
5376 DB CAGCAGCTAGACAGCAAAACCTCGAGGTGGAGAGAGATCTGTTTCAGTGTTTTATACACT 5435
1956 QY GlyHisAlaPheLeuAlaLysProProAlaAlaArgSerGlyLeuGluProAlaSerSer--- 1974
5436 DB TCTTCAGCTTTTCCAAAGTGGCAAGCCC-----CAGCTCATCTTTCAGTA 5480
1975 QY -----ProSerLysGlySerGluProArgProLeuValProValSerGlyHisAla 1992
5481 DB GTTTATTTCTGAGCTGGGAAAGATAAAGGCGCT---CCTCCAAATCCAGATATGAG 5534
1993 QY ThrIleAlaArgThrProAlaLysAsn-----LeuAlaProHisHisAlaSerProAsp 2010
5535 DB GAAGAGCTAAGGACACAGAGGGAAGACTCCATTACTGCAGCTAACTTCTATAGACGTGATC 5594
2011 QY ProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
5595 DB ATACCCGGCAAAATGCTCTCGACAGGATCGCAGGAACTGCTGCTCAAGATTCAGAC 5654
2031 QY PheSerIleGlnGluLeuLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
5655 DB TCTTCT-----AGTAGCTTATCTTCTCAGAGTATGAACACCTAGC 5696
2051 QY GluGlyValGluProValSerProValSerSerSerLeuThrHisAspLysGlyLeu 2070
5697 DB GATGCTATTGAGGTGATAGTCTCGCAGCTCAGCTGCGCCACCCCGAGGAACTGCGAG 5756
2071 QY ProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGln 2090
5757 DB ACCTATCAGCAGAGGTGTTAAGGCNAATCAAGCGGAAATGATCTTACAGACAAATAT 5816
2091 QY ProGlyProValLysLeuGlyGlyAlaAlaHisLeuProHisLeuArgProLeuPro 2110
5817 DB GAAGGACCA-----TTACATCATCATCGACCA----- 5843
2111 QY GluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGly----- 2125
5844 DB ---CAGCAGGAATCACCATCTCCCAACACAGCTGCCCTTCTTTCACAGCGAGAGGA 5900
2126 QY -----ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluVal 2142
5901 DB ATGGGGCAAGTGGCCAGGACCCATCGCTGATCAGACTGCTGATCACTGTCATAATT 5960
2143 QY IleThrGlnAspTyrThrArgHisHis-----ProGlnGlnLeuSerAla 2157
5961 DB ATCACACAAGATTTTGTAGAAATCAAGTTTCTTCGAGAGTCTCCCGACAG----- 6011
2158 QY ProLeuProAlaProLeuTyrSerPheProGlyAla-----SerCysProValLeuAsp 2175

Db 2172 GCAGAGCAGATGATGTAGATCAGCAGGAGCACAGTCTGTAAGAGGGTTCTGTTTGTGAT 2231
Qy 845 ProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSer 864
Db 2232 CCCCCCCTACCAAA-----GCTGACTCTGTGGACGTTGAAGTGGGTGCAGAA 2285
Qy 865 GluCysThrGluAlaGluGluGluProAlaLysGlyLysAspAlaGluAlaGlu 884
Db 2286 AACCATGTCATCTAAAGTGAAGGTGATAATACCAAGAAAGAGACTTGGATAGAGCC--- 2342
Qy 885 AlaThrAlaGluGlyAlaLeuLysAlaGluLysGlyGlySerGlyArgAlaThr 904
Db 2343 -----AGTGAGAGGTGGAACCTAGACATGAAGATTTGGTG 2378
Qy 905 ThrAlaLysSerSerGlyAla-----ProGln---AspSerAspSerSerAla 919
Db 2379 GTAGCTCAGCAATAATGCCAAAGGCCCGAGCCCGAGTCAGACATGATTCAGTGCC 2438
Qy 920 ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuSer 939
Db 2439 ACGTGACGCGCTGATGAG-----GATGTGATGGAGAGCCAGAGGCGAGAGATGTTT 2492
Qy 940 Pro-----ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSer 956
Db 2493 CCTGGACTCAAGGCTTCACTGTTAAACCCCACTGGATCTATACCTGCTC---TCATCT 2549
Qy 957 ProGlnLys-----ProLeuAspLeuLysGlnLysGlnArgAlaAlaAlaLeuPro 974
Db 2550 CCCTTAACCAATCCACTGGATCTGCCACAGCTTCAGCATCGAGCTGCTGTATCCCA 2609
Qy 975 Profile----- 976
Db 2610 CCAATGTATCTCTGCCACCCCATGTAAACATCAATGGAACCCAGTGGCGGTATGCT 2669
Qy 977 -----GlnValThrLysValHisGluProProArgGluAspAlaAlaProThr 992
Db 2670 CTCTACGAGCACATTAAGCAATGTCATGATGACAGCTCTCTGGAG----- 2717
Qy 993 LysProAlaProAlaProProProGlnAsnLeuGlnProGluSerAspAlaPro 1012
Db 2718 -----GAGCAGGCGAGAGACAGACATAGATTGGAA 2753
Qy 1013 GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProAlaAspLys 1032
Db 2754 TGTAGAAGTTCTACAGTCCATGTGGCAGATCCAAAGAGTCCA-----AACAGA 2801
Qy 1033 GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysThrProSerGly 1052
Db 2802 GAG-----TGG----- 2807
Qy 1053 LeuProPheProValProProArgGluValLysAlaSerProHisAlaProAspPro 1072
Db 2808 -----GAAGTCTCTCAGCTCTCTCACT----- 2831
Qy 1073 SerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla 1092
Db 2832 CAAGTGATACTAATCTCCCTGAAGGGGTTCCGGCTCCG-----ACAACT 2876
Qy 1093 ArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAla 1112
Db 2877 CGACCAACACGAGCCACGCGCCCTCTCATCCGTCATCCAAAACCCAGCTGGCTTCAGAA 2936
Qy 1113 LysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGly----- 1128
Db 2937 AAA---CCATCTTTATA-----ATGGGAGGTCTCCATCTCAGAGGAACACCGGCACT 2987
Qy 1129 ---MetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaPro---ValGly 1146
Db 2988 TATTTGACTTCTCATATCAGGCTTCTACACTCAAGAAACACCAAGCCGTCAGTGGGA 3047
Qy 1147 ProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGly 1166
Db 3048 TCTATCTCTCTTGAGCTGCCAGCGCAACAGGAATCTGCCAAATCAGCTACTTTGCCCTAC 3107

Qy 1167 ValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyVal 1186
Db 3108 ATCAAGCAGGAAGAAATTTTCTCCCCGAAGCCAAACCTCAACAACCTCAGGGTCTGTTGGTC 3167
Qy 1187 ProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGly 1206
Db 3168 ---AGGCGCCCAACATGAAGGTGTAGTCAGAGGTACCGCA---GGAGCCATACAGAAGGA 3221
Qy 1207 SerIleThrLysGlyLeuProSerThrArgValProSerAspSerAlaIleThrTyrArg 1226
Db 3222 AGTATTAACCTCGGGAACTCCACCAAGCAAAATTTAGTGAGAGCATTTCCATCCCTACGG 3281
Qy 1227 GlySerIleThrHisGlyThrProAla-----AspValLeu 1238
Db 3282 GGTCTCTATCACTCAGGCGACCCCGCTCTGCCAGACTGGCATACCAACAGAGGCTTTG 3341
Qy 1239 TyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGly 1258
Db 3342 GTCAAGGGGTCCATTTTCGAGAAATGCCCATTTGAAGACAGCAGTCTCT-----GAGAAAGGC 3395
Qy 1259 ArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysGlyHisValLeu 1278
Db 3396 AGAGAGAAAGCTGCATCCAAAGGCCATGTTATTTATGAAGGCCAAAGTGGACATATCTTG 3455
Qy 1279 SerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSer 1298
Db 3456 TCATATGATAATATTAAAGATGCC-----CGAGAAGGACTAGAGTCCA 3500
Qy 1299 GlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgVal 1318
Db 3501 AGAACAGCTCATGAATATCATGTTTA---AAGAGAAGCTATGAATCAGTGAAGGAATATA 3557
Qy 1319 GlyArgAlaIleSer-----SerAlaSerIleGluGlyLeuMetGly 1332
Db 3558 AGCAAGGATGTCATATGAGGAGTCTCTGTATCAGACCGTTAGAGGGCTGATATGC 3617
Qy 1333 ArgAlaIleProProGluArgHisSerProHis---HisLeuLysGlnGlnHisIle 1351
Db 3618 CGAGCATTAACCC---AGGGGAGTCTCTCATTTGACCTCAAGAAAGGACTGTATTG 3671
Qy 1352 ArgGlySerIleThrGlnGlyLeuProArgSerTyrValGluAlaGlnAspTyrLeu 1371
Db 3672 TCTGGCTCCATATGCGGGGACACCAAGAGCAACAACTGAAAGCTTTGAAGATGCGCTT 3731
Qy 1372 ArgArgGluAlaLysLeuLysArgGlyThrProProProProProProSerArg 1391
Db 3732 ---AAATATCCCAACAAATTAAGGGAAGT-----CCTCCCATACGA 3773
Qy 1392 AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHis 1411
Db 3774 GCATTTGAAGGTGCCATT-----ACCAAGGAAACCA---TAT 3809
Qy 1412 GluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGlu 1431
Db 3810 GATGGC---ATCACCACCATCAAGAAATGCGGCGTTCCATTTCATGAGATTCCAAGGCA 3866
Qy 1432 GluLeu-----ArgHisThrProGluLeuProLeuAlaProArgProLeu 1446
Db 3867 GATATTTTAACCTCAGGAAAGTCCGAAATCTCCAGAAAGTGTCCAGAGCACACCGCGCAT 3926
Qy 1447 LysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThr 1466
Db 3927 ATTGAGGGTTCATTTCCCGGGCACCACTAATAAGTTTGACAAACAC---TCAGGTCAA 3983
Qy 1467 GlySerLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProPro 1486
Db 3984 TCTGCCATCAAAACACATGTCAAATCTCTTAATCAGCGGGCTAGCAACTATCCCGTGA 4043
Qy 1487 ValHisProLeuAspValMetAlaAsp---AlaArgAlaLeuGluAlaCysTyrGlu 1505
Db 4044 ATGCCTCCCGTGGAAATTTGTGCCAGAGAAATANAAGTGTAGAACCGGGAATATGAG 4103


```
QY 2078 AspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGly 2097
Db 2232 AAGGCAATCAACGGGAATGATCTTACCAGCAATATCAAGGACCA-----6279
QY 2098 GlyGluAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSer 2117
Db 6280 -----TTACATCACTATCGACCA-----CAGCAGGAATCACCATCT 6315
QY 2118 ProLeuLeuGlnThrAlaProGly-----ValLysGlyHis 2129
Db 6316 CCCCAACAACAGCTGCCCTTCTTCCAGGAGAGGGAATGGGCAAGTCCAGGACC 6375
QY 2130 GlnArgValThrLeuAlaGlnHisLeuSerGluValIleThrGlnAspThrArg 2149
Db 6376 CATCGGCTGATCACATCTGCTGATCATCTGTCAAATATATCACAGATTTTGCTAGA 6435
QY 2150 HisHis-----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyr 2164
Db 6436 AATCAAGTTTCTCGCAGACTGCCAGCAG-----CCTCTACTTCTACATTCCAG 6486
QY 2165 SerPheProGlyAla-----SerCysProValLeuAspLeuArgProProSerAsp 2182
Db 6487 AACTCACTTCTGCTTGTGTATCTACACCTGTG-----AGGACTAAACATCAAC 6537
QY 2183 LeuTyrLeuProProProAsp-----HisGlyAlaPro-----AlaArgGly 2196
Db 6538 CGTTACAGCCAGAAATCCAGGCTCAGTCTGTCCATCATCAAGACCAAGGTTCAAGGTC 6597
QY 2197 SerProHis-----SerGluGlyGlyLysArgSerProGluProAsnLys 2211
Db 6598 TCTCCAGAAAATCTTGTGGACAAATCCAGGGAAGTAGGCTCGAAAATCCCCAGAGAGG 6657
QY 2212 ThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProProGluGlyMet 2231
Db 6658 AGTCACGTC-----TCTCCGAGCCCTACGAGCCCATCTCCCAACCCAG----- 6702
QY 2232 ThrGluProGlyHisSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
Db 6703 GTTCCGGTTGTGATGAAACAGGACAGCTTCTGCTCTGTCTCAGAGGGGGCAGAG 6762
QY 2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe 2271
Db 6763 CCTGCAGAGCAGGAATGATGCGCTCACAGGAGATATAAGCTACTTTCCTTCATTC 6822
QY 2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsn 2291
Db 6823 TTCACCAAGCTT---GAAAATACATCACCCATGTTAAATCAAGAAGCAGGAGATTTT 6879
QY 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThr 2311
Db 6880 CGTAGTTGAACCTCTCTGTGGGGTGTACTGTATATGCGAGCTGCTCAGCCAGGACT 6939
QY 2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
Db 6940 GAGATCTTTAATTCGCCAGGATTAATCTACCGAGCTCAGTTAGCTCTAGAGGCCATCT 6999
QY 2332 ValGlnGluHisAlaSerThrAsnMetGlyGluAlaIleIleArgLysAlaLeuMet 2351
Db 7000 TTTGCTGATCCTGCCAGT---AATCTGGGCTGGAAGACATTATTCAGGAAGGCTCTCATG 7056
QY 2352 GlyLysTyrAspGlnTyrGluGlu-----SerProProLeuSerAla 2365
Db 7057 GGAAGCTTGTATGCAAAAGTTGAGGATCATGGAGTTGTCTATGCCAGCTATGGGNGTA 7116
QY 2366 AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaMetProIleThrAla 2385
Db 7117 -----GTGCCCTGTGCTACTGCCAACACCTCAGTT-----GTGACCAAGT 7152
QY 2386 AlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly-----LysAla 2403
Db 7153 GGTGAGACACGAGAGGAGGAGGAGGCCATCATCTTTCAGGAGGAGTTTGCACCAACCA 7212
```

```
QY 2404 LysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro-----GlyLeu 2421
Db 7213 AAGCTGATCAGCAAGTCAAACAGCAGGAATCTAAGTCTCTCTATACCTGGGCAAGGCTAC 7272
QY 2422 AlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArg 2441
Db 7273 TTAGGAACCGAAGCGGCTCTTTCAGTCTCTCTGATCATTCAGAGGGGATTACCATAGG 7332
QY 2442 ArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThrPro 2461
Db 7333 CAGACGCCA-----GGTGGGCTGGGAAGACAGAGCCCTCTTCAACAGGCTCACTCAG 7386
QY 2462 PheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProPro 2481
Db 7387 TTTCTTATAAACCTCTGACTATGCGGATG-----CTCAGCAGTACTCCACCAACA 7437
QY 2482 ProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis-----HisAla 2497
Db 7438 CCGATTGTCATGCTCCCTCTCGGTGAACCAAGCAGCTCCTCACCACAGACAGGATC 7497
QY 2498 TrpAspGluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
Db 7498 TGGAGCGAGAGCTGCCCTCCCTCTCTCAGCACAGTACAGAGCCCTGTCGGATAGTAT 7557

RESULT 10
AAA60629
ID AAA60629 standard; cDNA; 7900 BP.
AC AAA60629;
XX
XX 24-OCT-2000 (first entry)
XX Human HNRCR encoding cDNA SEQ ID NO:19.
DE Human; HNRCR; nuclear receptor coreceptor; ss.
XX Homo sapiens.
OS
XX CN1250094-A.
PN
XX 12-APR-2000.
PD
XX 06-OCT-1998; 98CN-00120919.
PF
XX 06-OCT-1998; 98CN-00120919.
PR
XX (XINH-) XINHANGPU FUDAN GENE ENG CO LTD SHANGHA.
PA
XX Yu L, Tu Q, Zhao Y;
PI
XX WPI; 2000-400830/35.
DR
XX P-PSDB; AAB12453.
DR
XX Preparation of new human keron acceptor co-repressor coding series and the polypeptide.
XX Claim 1; Page 21-25; 58pp; Chinese.
PS
XX The present sequence encodes a human homologue of nuclear receptor coreceptor (HNRCR)
CC
XX Sequence 7900 BP; 2445 A; 1881 C; 1829 G; 1745 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1,97e-115 Length: 7900
Score: 4043.00 Matches: 1061
Percent Similarity: 54.21% Conservative: 330
Best Local Similarity: 41.35% Mismatches: 731
Query Match: 30.59% Indels: 444
DB: 3 Gaps: 104
US-09-522-753-5 (1-2517) x AAA60629 (1-7900)
```

Db 4237 -----ACCAAGGAAACCA---TATGATGC---ATCACCCACATC 4272
Qy LysGluAlaGlyArgSerIleHisGluLeuProArgGluGluLeu----- 1433
Db 4273 AAGAATGGGGCTTCCATTCATGAGATTCGAAGGCAAGATATTTAACTCAGGAAGT 4332
Qy ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGln 1453
Db 4333 CGAATACTCAGAGTGGTCCAGACACAGCGCGATATTTAGGGTTCATTTCCAG 4392
Qy GlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysHisAspVal 1473
Db 4393 GGCACACCAATAAGTTGACAAAC---TCAGTCAATCTGCCATCAACAAATGTC 4449
Qy ArgSerLeuIleGlySerProGlyArgThrPheProValHisProLeuAspValMet 1493
Db 4450 AAATCCTTAATCAGCGGCCCTAGCAAACTATCCCGTGGAAATGCTCCGCTGGAAATGTG 4509
Qy AlaAsp---AlaArgAlaLeuGluArgAlaCysTyrGlu-----GluSer 1507
Db 4510 CCAGAGACATAAAGTGTAGAACGGGGAATATAGGATGTGAAGCAGCGAGACC 4569
Qy LeuLysSerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaPro 1527
Db 4570 GTGGTTCGCGCACAGCTCAGTGTGAAGCTCTGGCCCTCCGTTCTTAGGTCCACA--- 4626
Qy ValIleValProGluLeuGlyLysProArgLysSerProLeuThrTyrGluAspHisGly 1547
Db 4627 -----CTGCATGAAGCTCCCAAGCACAACTAGCCCTCGGATTTATGATGACACAGT 4680
Qy Ala-----ProPheAlaGlyHisLeuProArgGlySerProValThrMet 1562
Db 4681 GCACGGAGGCCCTGTGAGTTATCAAAACACCATGTCCAGAGCTCACCATGATGAAC 4740
Qy ArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAsp 1582
Db 4741 AGAACTTCTGATGTACAATCTCT-----CCTAAACAAGTCTACCAATCATGAA 4788
Qy ArgLysLeuThrSerThrProArgGlu-----IleAlaLysSerProHisSer 1598
Db 4789 AGAAATTCGACATGACCCCTATCCAGAGGAAAGTATCCAGCGAAGTCTCCAGTGCCT 4848
Qy ThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyVal 1618
Db 4849 GGGGTGACCTGCTGTGACCCAC-----AGTCCGTTGTATCCCATCATCAGAGGACG 4902
Qy SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db 4903 ACTCAGCGCAGGTTTATTGGAGCCACCTGCCACGCAATGGATCCA---GCCATGCCT 4959
Qy ArgGlyIleProLeuAsp---AlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
Db 4960 TTTTCACAGGCTTTGGATCTCGACGCGCTGTACCTGTTTCAGACAGACTTTCACCA 5019
Qy AspProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAla 1677
Db 5020 ACTCCAGTTACCAAGTCAATCAGCTTTAC----- 5052
Qy AlaLeuGluAsn---ArgGlnThrIleAsnAspTyrIleThrSerGlnMetHis 1696
Db 5053 GCAATGGAGAACACAGACAGACAATCTTAAATGATTACATTACCTCACAACAGATGCAA 5112
Qy HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArg 1716
Db 5113 GTGAACCTTG-----CGTCCAGATGTGCCAGAGACTCTCCCCAAGA 5154
Qy GluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGln 1736
Db 5155 GAGCAGCCTTGGTCTCCCATACCCAGCA---ACGAGAGGAATCATTTGACCTGACCAAT 5211
Qy ValProHisLeuProValProThrProGlyThrProAlaThrAlaMetAsp 1756
Db 5211 ----- 5231

Db 5212 ATGCCTCCA---ACAAATTTAGTGCCTCATCCAGGGGGAACAAGCACTCTCTCCATGGAC 5268
Qy ArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSer 1775
Db 5269 AGAATCACTTATATTTCTGTGTACACAGATTACTTTCTCCAGCGGTACAACTCTGCT 5328
Qy ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db 5329 TCCATGTCTCCAGACACCCCAACACACCTT-----GCAGCTGTCTGCAAGTGTCTGAG 5379
Qy ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeu 1815
Db 5380 AGGAAACGGGACCGGAGCGGAGAGCGGAGCGGAGCGGAGCGG-----ATTGCT 5430
Qy ThrSerThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSer 1835
Db 5431 GCAGCTTCTCTCCACCTC-----TACTCGCGCCAGGCTCAGAACAG----- 5472
Qy GlySerSerGlySerSerGlyGlyGlySerSerArgProAlaSerHisSer 1855
Db 5473 -----CCTGGCCGACCTGCGCAGTCATGGA 5496
Qy HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArg 1874
Db 5497 TATGTTGCTGCTCCCTTCCCT---TCAGTAAGAACTCAGGAGACCATGTTTSCAACAGAGA 5553
Qy ProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLys 1894
Db 5554 CCAGTGTTTTCCAGGAACCAATGGAACCATGTGTAATCACCTTGGATTCACACTGCT 5613
Qy ProThrValLeuArgSerThrSerThrSerProVal-----ArgProAla 1910
Db 5614 CAGCTAGCAATCATGCACTGCTGCTGGGGCCCTTCAATAAGCAAGCCCTGCCAGCC 5673
Qy AlaThrPheProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPro 1930
Db 5674 TCCGTTTACAACTGCTGCGGATGCC---CTGGTCTCTTGTGATGCTGCGAGCTTCT 5730
Qy ThrLeuMetGluProValLeuLeuProLys-----GluAlaProArgValAla 1946
Db 5731 GCACCCAGATGATGTGTCCAAAACAAAGACAGTAAGCATGAAGCTGCCAGGTTAGAA 5790
Qy ArgProGluArgProArgAlaAsp----- 1954
Db 5791 GAAAATTTGAGAGCAGGTCAGCAGCAGTGTAGTGAACAGCAGCAGCTAGACAGAAAACC 5850
Qy 1955 -----ThrGlyHisAlaPheLeuAlaLys 1962
Db 5851 CTGGAGTGGAGAAGAGATCTGTTTACATGTTTATACACTTCTTTCAGCCTTTTCCAGTGC 5910
Qy ProProAlaArgSerGlyLeuGluProAlaSerSer-----ProSerLysGlySer 1979
Db 5911 AAGCCC-----CAGCCTCATCTTTCAGTAGTTTATCTAGGCTGGGAAA 5955
Qy GluProArgProLeuValProValSerGlyHisAlaThrIleAlaArgThrProAla 1999
Db 5956 GATAAAGGCT-----CCTCAAAATCCAGATATGAGGAAGAGCTTAAGACACAGAGG 6009
Qy LysAsn-----LeuAlaProHisHisAlaSerProAspProProAlaProAlaSer 2017
Db 6010 AAGACTACCATTAATCTGACGCTAACTTATAGACGTGATCATCACCGGCAAAATGCTCG 6069
Qy AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeu 2037
Db 6070 GACAAGATGCGAGGAACGTGGCTCTCAAAAGTTACACTTCT----- 6114
Qy LeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGlyGlyValGluProValSer 2057
Db 6115 ---AGTAGCTTATCTTCTCACAGGTATGAACACCATCAGCATGCTATTAGGTGATAAGT 6171
Qy ProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeu 2077
Db 6172 CCTGCCAGCTCACCTGCGCCACCCAGGAGAAAACCTGACACCTATCATGCCAGGAGTTGTT 6231

QY 751 IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPro 770
Db 2422 -----GTTGAAGCTGTCAAG----- 2436
QY 771 AlaThrLeuGlyAlaAspGlyProProGlyProProThrProProArgArgThrSer 790
Db 2437 -----CCGAGGAGGACAGTCTCTGAAATGTCTACTTCT 2469
QY 791 ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProPro 810
Db 2470 CGAGGAAACACAGAACCTCGGTTGAGCTTGAG-----CCACACGGAACACT 2517
QY 811 AlaPro-----ProSerProSer-----AlaProProValValProLysGluGlu 828
Db 2518 GCACCCAGTACATCTCCCTCTTAGCAGTTCCAGTACAAAACCCAGCTGAAGTAAAGT 2577
QY 829 GluGluThrAlaAlaAlaProProVal----- 837
Db 2578 GTGGAGACCCGAGTGAATGACAGCATCAGTCAGTCTGAGACAGCAGCAGATGGATGAT 2637
QY 838 -----GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu 851
Db 2638 CAGCAGAGCAGCAGTCTGAAGAGGGTCTGTTGTGATCCCCACCCGCTACCAAA--- 2694
QY 852 AlaValAspThrGlyLysAlaGluGluProProValLysSerGluCysThrGluGluAlaGlu 871
Db 2695 ---GCTGACTCTGGACGTTGAAGTGAAGTCCGAGGAGTCCGAGAAACCATGCATCTAAAGTTGAA 2751
QY 872 GluGlyProAlaLysGlyLysAspAlaGluAlaGluAlaGluAlaThrAlaGluGlyAlaLeu 891
Db 2752 GGTGATAATACCAAGAAAGAGACTTGGATAGAGCC----- 2787
QY 892 LysAlaGluLysGlyGluGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla 911
Db 2788 ---AGTGAGAAGTGAACCTAGAGATGAAGATTGGTGTGATGCTCAGCAAAATAATGCC 2844
QY 912 -----ProGln-----AspSerAspSerAlaThrCysSerAlaAspGluVal 926
Db 2845 CAAAGGCCCGAGCCCGCAGTCAGACATGATTCAGTCCAGTCCAGTCCAGCGCTGATGAG--- 2901
QY 927 AspGluAlaGluGlyAspLysAsnArgLeuLeuSerPro-----ArgProSer 943
Db 2902 ---GATGTGATGGAGCGCAGCAGCAGCAGAGAGTGTTCCTATGGAATCAAAAGCCTTCA 2958
QY 944 LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeu 961
Db 2959 CTGTTAAACCCCACTGATCTACTCTGTC---TCATCTCCGTTAAACCAAAATCCACTG 3015
QY 962 AspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIle----- 976
Db 3016 GATCTGCCACAGCTTCAGCATCAGCTGCTGTTATCCCAACCATGGTATCTCTGCACCCCA 3075
QY 977 ----- 3135
Db 3076 TGTAACATACCAATTGGAACCCCGAGTGGAGCTGCTCTTACACGAGCAGACATTAA 3135
QY 980 LysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaPro 999
Db 3136 GCAATGATGATGAGTCACTCTCGGAG----- 3162
QY 1000 ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro 1019
Db 3163 ---GAGCAGCGGAGCAGACAGACAGATAGATTGGAATGTAGAGTTCTTCAAGTCCA 3219
QY 1020 ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaGluAla 1039
Db 3220 TGTGGCACAATCCAAGAGTCCA-----AACAGAGAG----- 3249
QY 1040 GlnLysLeuProGlyAspProProCysTyrThrSerGlyLeuProPheProValProPro 1059
Db 3250 -----TGG----- 3252

QY 1060 ArgGluValIleLysAlaSerProHisAlaProAspProSerSerAlaPheSerTyrAlaPro 1079
Db 3253 ---GAAGTCTTCCAGCTGCTCCACAT-----CAATTGATAACTAATCTCCCT 3297
QY 1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
Db 3298 GAAGGCGTTCCGCTCCG-----ACAACTCGAACCAACCGAGCCACCGCCC 3342
QY 1100 ProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGlu 1119
Db 3343 CCTCTCATCCGTCATCCAAAACCCAGCTGGCTTCAGAAAA---CCATCTTTTATA--- 3396
QY 1120 ArgGlnIleGlyAlaIleSerGlnGly-----MetSerValGlnLeuHis 1134
Db 3397 ---ATGGGAGGCTCCATCTCACAGGGAACACACAGGCACTATTGTGACTTCTCTATAATCAG 3453
QY 1135 ValProTyrSerGluHisAlaLysAlaPro---ValGlyProValThrMetGlyLeuPro 1153
Db 3454 GCTTCTCAGCTACACAGAAACACCCAAAGCGCTCAGTAGGATCTATCTCTTGTGACTGCCA 3513
QY 1154 LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSer 1173
Db 3514 CGCAACACAGATCTCCAAATCAGTACTTGGCCCTACATCAAGCAGGAGAGATTTTCT 3573
QY 1174 ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSer 1193
Db 3574 CCGCAAGCCAAAACCTCACAACTGAGGCTGTGTTGTC---AGGCCCCAACATGAAGT 3630
QY 1194 ValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIlePro 1213
Db 3631 GTAGTCAGAGGTACCGCA---GGAGCCATACAAGAAAGATATACTCGGGAAGTCCA 3687
QY 1214 SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThr 1233
Db 3688 ACCGCAAAATTTCACTGGAGAGCATTCATCCCTAGGGGCTCTATCACTCAGGGCACC 3747
QY 1234 ProAla-----AspValLeuTyrLysGlyThrIleThrArg 1245
Db 3748 CCGGCTCTGCCAGACTGGCATACCAACAGAGGCTTGGTGAAGGGTCCATTTCGAGA 3807
QY 1246 IleIleGlyLysAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLys 1265
Db 3808 ATGCCCATTTGAAGACAGCAGTCTCT-----GAGAAAGGAGAGAGAGTGCATCCAAA 3861
QY 1266 GlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGlyGlyMetSer 1285
Db 3862 GGCATGTTATTTAGAGCAAAAGTGGACATATCTTGTTCATATATATTAAGAT 3921
QY 1286 ValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAla 1305
Db 3922 GCC-----CGAGAAGGAGCTAGGAGTCCAAGAACAGCTCATGAAATCAGT 3966
QY 1306 AlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer----- 1323
Db 3967 TTA---AAGAGAAGCTATGAATCAGTGAAGGAAATATAAAGCAAGGAGTGTCAATGAGG 4023
QY 1324 -----SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg 1339
Db 4024 GAGTCTCTGTATCAGCACCGTTAGAGGGGTGATATGCGGACATTAACC-----AGG 4077
QY 1340 HisSerProHis---HisLeuLysGlnHisHisIleArgGlySerIleThrGlnGly 1358
Db 4078 GGGAGTCTCTCATCTCAGCTCAAAAGAAAGGACTGTATTGTCTGCCTCCATAATGCGGGG 4137
QY 1359 IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLeu 1378
Db 4138 ACACCAAGAGCAACAACCTGAAGCTTTGAAGATGGCCTT---AAATATCCCAACAATAAT 4194
QY 1379 LysArgGlyThrProProProProProProProProProProProProProProProPro 1398
Db 4195 AAAGGGAAGT-----CCTCCCATAGAGCATTTGAGTGCATTT--- 4236
QY 1399 ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal 1418

Db 349 GAGTTCAGTCCCTGATTAATCGTCTCTCATCTTGAAGTGAAGTCAAGCATCAGCTT 408
Qy 54 SerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGln 73
Db 409 TTGCAGCAACAGCAGCAGCAACAGCTTCGAAGCGACCTTCCTTCCTTCCTTCAGAAATTCAC 468
Qy 74 ProGlyAsnGluArgSerGlnGlnLeuLeuHisLeuArgProGluSerHisSerTyrLeuPro 93
Db 469 CCAGGTTCTGACAGCGCTCAGAA-----AGAGAACTAGTTATGAACCGTTTCATCCA 522
Qy 94 GluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeu 113
Db 523 GGCCCATCCCGAGTGCATCATGATCTCATGATCGAGCGACCATCGTCTGGACAGGTT 582
Qy 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 583 TCTGATTTCTATTTCAGCGTGTCAAGTCTGCGGTTTTCGCTTTAGTGCACCCGCTGCCA 642
Qy 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 643 -----GAAGGGCTGAGGGCT---TCTGCAGATGCTTAAGAGGATCCAGCATTCGGA 690
Qy 143 GlyLysLeuGlu---ProValSerProProSerProProHisThrAspProGluLeuGlu 161
Db 691 GGCACACATGAAGCTCCATCTCTCCATTTTCGGGGCAACCATGTGGAGATGATCAAAAT 750
Qy 162 LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArg 181
Db 751 GCTTCACTTCAAACTCTCAAGGAAGAGTAAATACAGAGTATGGATCGTGTAGATCGA 810
Qy 182 GluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGlu 201
Db 811 GAAATTCAGAAAGTAGAACAGCATCTTAAACTGAAAGAAAGAAACAAACAGCTTGAA 870
Qy 202 GluGluAlaAlaLysProProGluProGluLysProValSerProProIleGluSer 221
Db 871 GAAGGCGACATAAATCTCTGAGCTGAGAGCCCGTGTCCCTCTCTCTGTGGAGCAG 930
Qy 222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAla 241
Db 931 AAACACGCGAGTATGTCCAAATTTATTTATGATGAGAATCGGAAAGAGAGAGCT 990
Qy 242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
Db 991 CATAAATTTTGAAGTCTTGCCCAAGAGTTGAAGTGGCACTGTATATACAGCCATCA 1050
Qy 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
Db 1051 GATACCAAGGTGTACCATGAGACATCAAGACAAACCCAGGTGATGAGGAAACAACTCATT 1110
Qy 282 LeuTyrPheLysArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg 301
Db 1111 TTATTTTTTAAAGAGAAATCATGCAAGAAACAAAGGAGCAACAAATAATCTCCAGCGT 1170
Qy 302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArg 321
Db 1171 TATGATCAGCTCATGGAGCGCATGGGAGAAAAGTGGACAGATAGAAAATATATCTCCG 1230
Qy 322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341
Db 1231 AGGAAAGCTTAAAGAAACAAACAAAGGAATCTATGAAAGCAGTTTCCAGAAATTCGA 1290
Qy 342 LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu 361
Db 1291 AAACAAAGAGAACACGCAAGAAAGATTTTCAG---CGAGTTGGCGAGAGGAGCTGTCTT 1347
Qy 362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu 381
Db 1348 TCAGCCACCATTTGCTAGAGTACGATGAGATTTCTGAAATTTATGATGGCTCTCTGAG 1407
Qy 382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp 401

Db 1408 CAGGAGATAATGAGAAACAAATGCGGAGCTCTCTGTGATTCACCTATGATTTTGTAT 1467
Qy 402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
Db 1468 GCAGAACAAAGACGATCAAGTTCAATGATCAATGATGGCTTATGAGGACCCCTATGAA 1527
Qy 422 ValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPheArg 441
Db 1528 GTGTATAAGATAGGACGATTTATGATGTTTGGACGACCATGAAAGAGGATCTTTAAG 1587
Qy 442 GluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLys 461
Db 1588 GACAAAGTTTATCCAGCATCCAAAAAATTTGGACTAATTCATCATCTACTTGGAGAGAG 1647
Qy 462 ThrValAlaGluCysValLeuTyrTyrLeuThrLysLeuAsnGluAsnTyrLysSer 481
Db 1648 AGTGTTCCTGATGTGTTTGTATTACTATTAAACCAAGAAAATGAGAAATATAAGCC 1707
Qy 482 LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
Db 1708 CTCTCAGAAAGAAATATGGAAACCGCAGAGCGAGAACCCAGCAATTCGTCGACCTCG 1767
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlnGluLys 520
Db 1768 CAAGAGAAAAGTACAGAAAAGAGAG-----GATAAAGCAGAAAACAGAAAAA 1821
Qy 521 AspGluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsn 540
Db 1822 AAAGAAAGAAAGAAAGATGAAGAGGAAAAGATGAAGAAAGAAAGAAAGAAAGAAAT 1881
Qy 541 AspLysGluAspLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu 560
Db 1882 ACCAAGAA-----AAGGACAGATAGATGTTACAGCA--GAAGAACTGAGGAA 1929
Qy 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
Db 1930 AGAGAGCAAGCCACACCCCGGGCGGAAAGACTGCCAACAGTCAGGCGCGCGTAAGGC 1989
Qy 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1990 CGATCACCAGGTCCATGACAAACAGAGCTGCGAGCTGCCAGTCCGCGCGCGTAAGGC 2049
Qy 601 Ser-----AlaGluLeuAlaSerMetGluLeuAsn 610
Db 2050 ACTGAAGAGCCCCCACCCTCTGCCACCGCCACAGACCCATTTCTACAGAGCTGTG 2109
Qy 611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuGluHis 630
Db 2110 GAGACCTCTCGATGGACAGAAAGAAATGGAAGTTGCTTAAAGAGGTCTAGTAGAACAT 2169
Qy 631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
Db 2170 GGTCTGTAAGTGGCAGCAATTTGCTAAATGGTGGGAAAGGAAAGTGAAGCTCAATGAAA 2229
Qy 651 AsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnHisLys 670
Db 2230 AACTTCTATTTTAACTATAAAGCGCAGACAACTCTTGACAACTCTTACAGAGCATAAA 2289
Qy 671 LeuLysMetLysLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaAla 690
Db 2290 CAGAAACCTTCAGAAACCTCTGTAAGAGCAGATGTGTCTCAATGTGAAGTGTGCT 2349
Qy 691 SerGluGluAlaAlaPheProValValGluAspGluMetGluAlaSerGlyVal 710
Db 2350 TCCACTGTTCTGCT-----CAGGAGGATGAGATATTGAAGCTCC----- 2391
Qy 711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 2392 -----AATGAGAGAGAAATCCAGAGACAGCGAA----- 2421
Qy 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
Db 2421 ----- 2421

Db :||:||||| ||| :||: ||||:||||:||||:||||:||||: 4369
Qy AAAATTTGACAACAAC---TCAGGTCAATCAGCTATCAAAACAATGTGAAGTCTTAAT
Db eGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAsp---Al 1496
Db :||:||||| :||: ||||:||||:||||:||||: 4423
Qy CACAGGCGCTAGCAACTA-----CCCGTGGAAATCTGGAATTTGACAGAAACAT
Db aArgAlaLeuGluArgAlaCysTyrGlu-----GluSerLeuLysSerAr 1511
Db :||:||||| :||: ||||:||||:||||:||||: 4483
Qy AAAAGTAGTAGACGGGAAATATAGGATGTGAAGCAGGCGCAGTCGCGAGCCG
Db gProGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValPr 1531
Db :||:||||| :||: ||||:||||:||||:||||: 4534
Qy GCACAGCTCAGTGTGAGCTCTGGCCCTCGTCTCAGGTCTACA-----CTTCA
Db oGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla----- 1548
Db :||:||||| :||: ||||:||||:||||:||||: 4594
Qy CGAAGCTCCCAAGACACAGCTGAGCCGCGGACTCTATGATGACAGCAGCGCTCCGAGGAC
Db :||:||||| :||: ||||:||||:||||:||||: 1566
Qy :||:||||| :||: ||||:||||:||||:||||: 4646
Db :||:||||| :||: ||||:||||:||||:||||: 1585
Qy :||:||||| :||: ||||:||||:||||:||||: 4693
Db :||:||||| :||: ||||:||||:||||:||||: 1601
Qy :||:||||| :||: ||||:||||:||||:||||: 4753
Db :||:||||| :||: ||||:||||:||||:||||: 1621
Qy :||:||||| :||: ||||:||||:||||:||||: 4807
Db :||:||||| :||: ||||:||||:||||:||||: 1641
Qy :||:||||| :||: ||||:||||:||||:||||: 4864
Db :||:||||| :||: ||||:||||:||||:||||: 1661
Qy :||:||||| :||: ||||:||||:||||:||||: 4924
Db :||:||||| :||: ||||:||||:||||:||||: 1681
Qy :||:||||| :||: ||||:||||:||||:||||: 4957
Db :||:||||| :||: ||||:||||:||||:||||: 1700
Qy :||:||||| :||: ||||:||||:||||:||||: 4958
Db :||:||||| :||: ||||:||||:||||:||||: 5015
Qy :||:||||| :||: ||||:||||:||||:||||: 1720
Db :||:||||| :||: ||||:||||:||||:||||: 5059
Qy :||:||||| :||: ||||:||||:||||:||||: 1740
Db :||:||||| :||: ||||:||||:||||:||||: 5114
Qy :||:||||| :||: ||||:||||:||||:||||: 1760
Db :||:||||| :||: ||||:||||:||||:||||: 5173
Qy :||:||||| :||: ||||:||||:||||:||||: 1779
Db :||:||||| :||: ||||:||||:||||:||||: 5233
Qy :||:||||| :||: ||||:||||:||||:||||: 1799
Db :||:||||| :||: ||||:||||:||||:||||: 5284
Qy :||:||||| :||: ||||:||||:||||:||||: 1819

Db 5285 ACGGAAAGGAGAGGAGCGGAAACGTGAGCCGCGGTGAGCGTGAACGTGAACCGGA 5344
Qy rThrValGluHisAlaPro-----lleTpaArgProGlyThrGluGlnSerSerGl 1836
Db :||:||||| :||: ||||:||||:||||:||||: 5345
Qy AAGATCGCTGCTGCTCCGCTGACCTTACCTACGACCAAGTTTCAGAACAG----- 5396
Db ySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaSerHisSerHi 1856
Db :||:||||| :||: ||||:||||:||||:||||: 5397
Qy :||:||||| :||: ||||:||||:||||:||||: 1875
Db :||:||||| :||: ||||:||||:||||:||||: 5423
Qy :||:||||| :||: ||||:||||:||||:||||: 1875
Db :||:||||| :||: ||||:||||:||||:||||: 5479
Qy :||:||||| :||: ||||:||||:||||:||||: 1895
Db :||:||||| :||: ||||:||||:||||:||||: 5480
Qy :||:||||| :||: ||||:||||:||||:||||: 1895
Db :||:||||| :||: ||||:||||:||||:||||: 5540
Qy :||:||||| :||: ||||:||||:||||:||||: 1911
Db :||:||||| :||: ||||:||||:||||:||||: 5540
Qy :||:||||| :||: ||||:||||:||||:||||: 1911
Db :||:||||| :||: ||||:||||:||||:||||: 5599
Qy :||:||||| :||: ||||:||||:||||:||||: 1931
Db :||:||||| :||: ||||:||||:||||:||||: 5600
Qy :||:||||| :||: ||||:||||:||||:||||: 5656
Db :||:||||| :||: ||||:||||:||||:||||: 1944
Qy :||:||||| :||: ||||:||||:||||:||||: 5657
Db :||:||||| :||: ||||:||||:||||:||||: 5716
Qy :||:||||| :||: ||||:||||:||||:||||: 1945
Db :||:||||| :||: ||||:||||:||||:||||: 5717
Qy :||:||||| :||: ||||:||||:||||:||||: 1958
Db :||:||||| :||: ||||:||||:||||:||||: 5776
Qy :||:||||| :||: ||||:||||:||||:||||: 1978
Db :||:||||| :||: ||||:||||:||||:||||: 5836
Qy :||:||||| :||: ||||:||||:||||:||||: 1987
Db :||:||||| :||: ||||:||||:||||:||||: 5837
Qy :||:||||| :||: ||||:||||:||||:||||: 2005
Db :||:||||| :||: ||||:||||:||||:||||: 5897
Qy :||:||||| :||: ||||:||||:||||:||||: 2025
Db :||:||||| :||: ||||:||||:||||:||||: 5957
Qy :||:||||| :||: ||||:||||:||||:||||: 2045
Db :||:||||| :||: ||||:||||:||||:||||: 6016
Qy :||:||||| :||: ||||:||||:||||:||||: 2045
Db :||:||||| :||: ||||:||||:||||:||||: 6058
Qy :||:||||| :||: ||||:||||:||||:||||: 2065
Db :||:||||| :||: ||||:||||:||||:||||: 6118
Qy :||:||||| :||: ||||:||||:||||:||||: 2085
Db :||:||||| :||: ||||:||||:||||:||||: 6178
Qy :||:||||| :||: ||||:||||:||||:||||: 2105
Db :||:||||| :||: ||||:||||:||||:||||: 6179
Qy :||:||||| :||: ||||:||||:||||:||||: 2105
Db :||:||||| :||: ||||:||||:||||:||||: 6211
Qy :||:||||| :||: ||||:||||:||||:||||: 2120
Db :||:||||| :||: ||||:||||:||||:||||: 6271
Qy :||:||||| :||: ||||:||||:||||:||||: 2138
Db :||:||||| :||: ||||:||||:||||:||||: 6331

Db 399 GGCCTTCCCGGTGGACCATGCTCGCTGGAGTCCAAAGCGGCTCGCCCTGGAGCAGGTT 458
Qy 114 -ProAspProLeu-----LeuArgProSerProLeuLeuAlaThrGlyGlnPr 129
Db 459 TCCGACTCCCACTTCCAGCGCATCAGTGTGCTGCGCTCTCCCTTTGGTGACACG-CTGCC 517
Qy 129 oAlaGly-----SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuG1 146
Db 518 AGAAGGACTGAGTCTTCTGCCAATGCTAAGAGGATCCGGCATTTGGAGTCAAAACATGA 577
Qy 146 u----ProValSerProProSerProProHisThrAspProGluLeuGluLeuValProPr 165
Db 578 AGCTCTCTCTCCCTCTCTGGGCAGCCATCGGAGATGATCAGAAATGCTCCACCTTC 637
Qy 165 oArgLeuSerLysGluLeuLeuLeuGlnAsnMetAspArgValAspArgGluLeuThrMe 185
Db 638 AAAACTCTCAAGAGAGCTGATACAGAGCATGGATCGTGTCTGATCGAGAAATTCGAA 697
Qy 185 tValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluAlaA1 205
Db 698 AGTAGAACAGCAGATCCTTAACTGAAAGAAAGACAGCAACAGCTCGAAGAAAGCTGC 757
Qy 205 aLysProProGluProGluLysProValSerProProProIleGluSerLysHisArgSe 225
Db 758 TAAACCCCGCAGAGCTCAGAAAGCTGTGTCCTCTCCCTCGGTGGAGCAGAACCGAAG 817
Qy 225 rLeuValGlnIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLe 245
Db 818 TATTGTCAAATCATTTATGACGAGATCGGAAAGACAGAGAGCTCATATAAATATT 877
Qy 245 uGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgG1 265
Db 878 TGAAGGTCTTGGCCCAAGTGTGAACCTGCGCTCTACACAGCCGTCAGATACCAAGGT 937
Qy 265 nTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLy 285
Db 938 GTACCACGAGACATCAAGACAAACAGGTGATGAGGAAACCTCATTTTATTTTAA 997
Qy 285 sArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLe 305
Db 998 AAGAAGAAATCATGCAAGAAACAAAGGAAACAAACAAATCTGCCAACGTTATGATCAGCT 1057
Qy 305 uMetGluAlaLeuGluLysValGluArgLleGluAsnProArgArgArgAlaLy 325
Db 1058 CATGGAAGCATGGGAGAAAGAGTGACAGAAATAGAAAATCTCCTCGAGGAAAGCTAA 1117
Qy 325 sGluSerLysValArgGluTyrTyrGluLysGlnPheProGluLeuLeuArgLysGlnArgG1 345
Db 1118 AGAAGCAAAACAAAGGAATACTATGAAGACAGTTTCCAGAAATTCGAAACAAAGAGA 1177
Qy 345 uLeuGlnGluArgMetGlnSerArgValGlyGlnArgLysGlySerGlyLeuSerMetSerAl 365
Db 1178 ACAGCAAGAAAGATTTTCAG---CGAGTTGTGTGAGGGAGCTGCTTTTCAGCCACCAT 1234
Qy 365 aAlaArgSerGluHisGluValSerGluIleLeuAspGlyLeuSerGluGlnGluAsnLe 385
Db 1235 TGCTAGAGTGGAGCATGAGATTTCTGAATTTATGTGCTTCTGAACAGGAGAAATAA 1294
Qy 385 uGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnG1 405
Db 1295 TGAGAAGCAATGGTCAGCTTCTGTGATTTCCACCTATGATGTTGATGTCAGAACAAAG 1354
Qy 405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425
Db 1355 AAGGTCAAATTCATCAATATGATGAGTGGCTGATGGAGGATCCCAATGAAGTTTATAAGA 1414
Qy 425 pArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMe 445
Db 1415 CAGACAGTTTATGATGATGTTGAGCTGACCATGAAAGAGAGATCTTTAAGGACAGATTAT 1474
Qy 445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaG1 465

Db 1475 CCAGCATCCAAAAAATCTTGGACTAAATTCATATTCGAAAGGAAGAGTGTTCCTGA 1534
Qy 465 uCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgAr 485
Db 1535 TTGTGTTTATATATCTATTTAAACCAAGAAATAGAAATTAAGGCTCTCGTGAGAAG 1594
Qy 485 gSerTyr---ArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnG1 504
Db 1595 GAATATGGAANAACGACAGGACAGAAATCAGCAGATTCGCCGCTCCACAGAAGAAAA 1654
Qy 504 nGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysG1 524
Db 1655 AGTAGAGAAAG-----GAAGAGGATAAAGCAGAAAAAAC 1690
Qy 524 uLysGluLysGluAlaGluLys-----GluGluGluLysProGluValGluAsnAspLy 542
Db 1691 AGAGAAAAGGAGAGAAAGAGATGATGAAGAAAAAGATGATGAAGAGACTCTAA 1750
Qy 542 sGluAspLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysG1 562
Db 1751 AGAAACAACCAAGAAAGACAGACGGAAGCCACAGACAGAAACCTGAAGAAAGAGA 1810
Qy 562 uAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgI1 582
Db 1811 GCAGGTCACTCCAAGGGGGGAAAGACTGTCTAAACAGCAAGCGCGGGAAGGCGCGGT 1870
Qy 582 eThrArgSerMetAlaAsnGlu-----AlaAsnSerGluGluAlaIleThrProG1 599
Db 1871 CACAGGTCTGATCAAGTGAAGCTGAGCTGCTCAATGCTGCTGCTCAGCCACTGAGGA 1930
Qy 599 nGlnSerAlaGluLeu-----AlaSerMetGluLeuAsnGluSerSe 613
Db 1931 GCCCCGCCACCTCGCCGCCACACAGAGCCCATTTCTACAGAACCTGTGTGAGACTTC 1990
Qy 613 rArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAs 633
Db 1991 AAGATGACAGAGAAAGAAATGGAAGTTGCTAAAGAGGCTGTGTAGAACATGTCGTAA 2050
Qy 633 nTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTy 653
Db 2051 CTGGGACGACATTTGCTAAATGTGGGAACATAAAGTGAAGCCAGCGGCAAAACTTCTA 2110
Qy 653 rPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMe 673
Db 2111 TTTTAACTATAAAGACGCGCATTAATCTTGACACCTTTTCGACCAACATAAAGAGAGC 2170
Qy 673 tGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaSerGluG1 693
Db 2171 TTCACGGAACCCGTCGAGGAACGAGATGTATCTCAGTGTGAAGTGTGCTTCCACTGT 2230
Qy 693 uAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAs 713
Db 2231 TTCTGCC-----CAAGAGGATGAAGATATTGAAGCCTCA-----AA 2266
Qy 713 nGluGluGluMetValGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProAr 733
Db 2267 TGAGGAGAAATCCAGAACATAGTAGAGT----- 2297
Qy 733 gGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSe 753
Db 2298 -----GCTGAAAATAGTTCTGATACAGAAAAGTGTCCCTC 2332
Qy 753 rProHis---ThrGluAlaAlaLys-----AspThrGlyGlnAsnGlyProLysPr 769
Db 2333 TCCTTCAACAGTTGAAGTCCCAAGTCCAGTGAAGACAGCAGCATGAAATGCT----- 2384
Qy 769 oProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProArgArgTh 789
Db 2385 ----- 2386
Qy 789 rSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProPr 809
Db 2387 TTCTCGAGGAAACACCGAGGCTGTGGCTGAGCTTTGAGGCCACCACCTGACCCCTG 2446

3099 GGAGCCAGCAGCAGCCAGCGTCTCGAGGTCACCTCCACCTCCACCGCTTCCGCCAGC 3158
1910 aAlaThrProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPr 1930
3159 TGGCATTTCCTCCACCTGCGCACCTGCGCACCTGCGCGGCGACCTCGATGGGGTCTACCC 3218
1930 oThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAArgProGluAr 1950
3219 TACCCTCATGGAGCCGCTTGTGTCAGAGAGGCCCCCGGGTCCGCCGCGCAGAGCG 3278
1950 gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGl 1970
3279 GCCCGAGCAGACACCGCGCATGCCCTTCTCGCAAGACCCCGCGCTCGGGCTGGA 3338
1970 uProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGl 1990
3339 GCCCGCTCTCTCCCGCAGCAGGGCTCGGAGCCCGCGCCCTAGTGGCTCTCTCTGG 3398
1990 yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAlaSerProAs 2010
3399 CCAGCCACCATCGCCGCGCACCTCGCAAGAACCTCGCACCTCACACGCCAGCCGGA 3458
2010 pProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
3459 CCGCGCGCGCCACCTGCCCTCGGCCTCGACCGCGCACCGGAAAGACTCAAAAGTAAAC 3518
2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
3519 CTTTTTCATCCAGAACTGGAATCTCGTTCTCTGGGTACACCGGAGAGCTACAGCCC 3578
2050 oGluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLe 2070
3579 CGAAGGGGTGAGCCCTGAGCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGCT 3638
2070 uProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGl 2090
3639 CCCCAAGACACTGGAAGAGCTGACAGAGCCACTCGAGGGGGAGCTGCGGGCCCAAGCA 3698
2090 nProGlyProValLysLeuGlyGlyAlaAlaHisLeuProHisLeuArgProLeuPr 2110
3699 GCCAGGCCCTGAGCTTGGCGGGAGGCGCGCCACCTCCACACTCGCGCGCGCTGCC 3758
2110 oGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGl 2130
3759 TGAGAGCCAGCCCTCGTCCAGCCGCTGCTCCAGACCGCCCGCAGGGGTCAAAGGTCAACA 3818
2130 nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi 2150
3819 GCGGGTGGTCAACCCTGGCCCGACACATCAGTGAAGTTCATCACAGGAGCTACACCCGGCA 3878
2150 sHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe 2170
3879 CCACCCACAGCAGCTAGCGCACCCCTGCGCGCCCGCCCTCTACTCTCTCTCTGGGGCCAG 3938
2170 rCysProValLeuAspLeuArgProProSerAspLeuTyrLeuProProProAspHi 2190
3939 CTGCCCCGTCTGGACCTCGCGCGCCACCCAGTGACCTTACTCTCCCGCCCCCGGACCA 3998
2190 sGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAs 2210
3999 TGGTGGCCCCCGCGTGGCTCCCCCAGCAGGAGGGGCAAGAGGTCTCCAGAGCCAA 4058
2210 nLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProProGluGl 2230
4059 CAAGACGTCTGGTGGTGGTGGAGCGGTATTGAACCTGTGTCTCCACCGGAGGG 4118
2230 yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGl 2250
4119 CATGACGAGCCAGGGCAGCTCCCGAGTGTGTGTACCGCTGTGTACCGGGATGGGA 4178
2250 uGlnThrGluProSerArgMetGlyLysSerProGlyAsnThrSerGlnProAl 2270
4179 ACAGACGAGCCAGCAGGATGGGCTCCAAAGTCTCCAGGGCAACACCGAGCCGCGCAGC 4238

2270 aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIl 2290
4239 CTTCTTCAGAGCTGACCCAGAGCAACTCCGCCATGGTCAAGTCCCAAGAAGCAGAGAT 4298
2290 eAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGl 2310
4299 CAACAAGAAGCTGAACACCCACCAACCGAATGAGCCTGAATACAAATATCAGCCAGCCTGG 4358
2310 yThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGl 2330
4359 GACGGAGATCTTCAATATGCCCGCATCACCGAAGAGCCTTATGACCTATAGAGCCA 4418
2330 nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLe 2350
4419 GCGGTGTCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGCAC 4478
2350 uMetGlyLysTyrAspGlnTyrGluSerProProLeuSerAlaAsnAlaPheAsnPr 2370
4479 CATGGGTAAATATGACAGTGGGAAGAGTCCCGCGCTCAGCGCCAATGCTTTTAAACC 4538
2370 oLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSe 2390
4539 TCTGAATGTCAGTGCAGCCTGCCCTGCTATGCCCATTAACCGCTGTGACGAGCGAG 4598
2390 rAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSe 2410
4599 TGACCAACACATCACCTCGCCAGTTCGCGGGGAGGAGCCCAAGGTCTCTGGCAGACCCAG 4658
2410 rSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVa 2430
4659 CAGCCGAAAGCCCAAGTCCCGCGCCCTGGCATCTGGGGAGCCGCGCACCTCTGT 4718
2430 lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
4719 CTCCTCAGTGCACCTCGAGGAGAGCTGCAACCCCGGAGCCGCTCACCACCCGCGTGTG 4778
2450 pGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetAr 2470
4779 GGAGGACAGCCCTCGTCCGAGGTTCCAGCCATTCCCTACACCCCTGATCATCGC 4838
2470 gLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyPr 2490
4839 GCTGACGAGGGGTGTATGGCTTCCCGCCACCCCGGGCTCCCGCGGAGCGGGCC 4898
2490 oLeuAlaGlyProHisAlaAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTy 2510
4899 CTTGCTGGCCCCCACCACGCTTGGAGCAGGAGGCCAAGCCACTGCTCTGCTCGCAGTA 4958
2510 rGluThrLeuSerAspSerGlu 2517
4959 CGAGACACTCTCCGACAGCGAG 4980

RESULT 7

ADC35130

ID ADC35130 standard; cDNA; 2930 BP.

XX

AC ADC35130;

XX

DT 18-DEC-2003 (first entry)

XX

DE Human breast cancer antigen polynucleotide seq id 14.

XX

KW breast cancer; breast cancer diagnosis; breast cancer antigen; gene; ss.

XX

OS Homo sapiens.

XX

PN US2003108888-A1.

XX

PD 12-JUN-2003.

XX

PF 15-MAY-2002; 2002US-00146473.

XX

QY 1170 uGlnLeuSerProArgGlyGlnAlaGlyProGluSerLeuGlyValProThrAlaG1 1190
DB 939 GCAGCTGTCCCAACGGGGCAGGCTGGGCCACCGAGAGCCTGGGGTCCACAGCCCA 998
QY 1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy 1210
DB 999 GGAGGGTCCGTCTGAGAGGACAGCTCTGGGCTCAGTTCCGGGCGGAAGCATCACAA 1058
QY 1210 sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
DB 1059 AGGCATTCCAGCACACGGGTGCCCTCGGACAGCGCCATCACATACCGGGTCCATCAC 1118
QY 1230 rHisGlyThrProAlaAspValLeuTyrGlyThrIleThrArgIleIleGlyGluAs 1250
DB 1119 CCACGGCACCCAGCTGACGTCTGTACAAAGGACCATCACAGATCATTCGGCAGGA 1178
QY 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy 1270
DB 1179 CAGCCCGAGTCCGTGGACCGCGCGGAGGAGACAGCTGCCCAAGGGCCACGTCTA 1238
QY 1270 rGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe 1290
DB 1239 CGAAGGCAAGAGGCCACGCTTGTCTATGAGGTGGCATGTCTGTGACCCAGTGCTC 1298
QY 1290 rLysGluAspGlyArgSerSerGlyProHisGluThrAlaAlaProLysArgTh 1310
DB 1299 CAAGGAGGAGCGGAGAGCAGCTCAGGACCCGCCATGACAGCGGCCGCCCAAGCGCAC 1358
QY 1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
DB 1359 CTATGACATGATGAGGCGCGGTGGGACAGAGCCATCTCTCAGCCAGCATCAAGGTCT 1418
QY 1330 uMetGlyArgAlaIleProProGluArgHisSerProHisIleLeuLysGluGlnHisH 1350
DB 1419 CATGGCGGTGCCATCCCGCGGAGCGACACAGCCCCCACCACCTCAAGAGAGAGACCA 1478
QY 1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTy 1370
DB 1479 CATCCGGGGTCCATCACAAAGGATCCCTCGTCTAGTGGAGGACAGGAGGACTA 1538
QY 1370 rLeuArgArgGluAlaLysLeuLysArgGluGlyThrProProProProProSe 1390
DB 1539 CCTGGCTCGGAGGCCAAGCTCTAAAGCGGAGGAGCGCCCTCCGCCGCCACCGCCCTC 1598
QY 1390 rArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
DB 1599 ACGGACCTGACCGAGGCTTAAAGACGAGGCTCTGGGCCCTGAGCTGAGGAGCGGC 1658
QY 1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
DB 1659 CCATGAGGCTGTGGTGGCCACGGTGAAGGAGCGGCGCTCCATCCATGAGATCCCGCG 1718
QY 1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySe 1450
DB 1719 CGAGAGCTGGGACACCGCGAGCTGCCCTCGCCCGCGCGCTCAAGAGGAGGCTC 1778
QY 1450 rIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLy 1470
DB 1779 CATACGAGGACACCCCGCTCAAGTACGACACCGCGCGCTCCACCCTGCTCCAAAAA 1838
QY 1470 eHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLe 1490
DB 1839 GCACGACGTACGCTCCCTCATCGGACGCCCGCGGAGCGTTCACCCGCTGCACCCGCT 1898
QY 1490 uAspValMetAlaAspAlaArgAlaLeuLysArgAlaCysTyrGluGluSerLeuLysSe 1510
DB 1899 GAGTGTATGGCGGACGCGCGGCTCAAGTACGACCGCGCTCCAGGAGAGCCCTGAAGAG 1958
QY 1510 rArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaProValIleVa 1530
DB 1959 CCGGCCAGGAGCGGACGAGCTCGGGGGCTCCATTGGCGCGCGCCCGCTCATTTGT 2018
QY 1530 lProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPh 1550

DB 2019 GCCTGAGCTGGGTAAAGCCGCGGACAGCCCTGACCTATGAGGACCAACGGGGCACCCCTT 2078
QY 1550 eAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuG1 1570
DB 2079 TGCCGGCCACCTCCACAGGTTCCGCCGTGACCATGCGGAGGCCACGCGCGCTGCA 2138
QY 1570 nGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProAr 1590
DB 2139 GGAGGAGCAGCTTCGTCCAGCAAGGCATCCAGGACCGAAGCTGAGCGTCCAGCCCTCG 2198
QY 1590 gGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPr 1610
DB 2199 TGAGATCGCAAGTCCCGGCACAGCACCGTGGCCGAGCACCCACCCCATCTCGCC 2258
QY 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProle 1630
DB 2259 CTATGAGCACCTGCTTCGGGGGTGAGTGGCGTGACCTGTATCGCAGCCCATCCCT 2318
QY 1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTy 1650
DB 2319 GGCTTCGACCCACCTCCATACCCCGGGCATCCCTCTGGACGAGCGCTGCTACTTA 2378
QY 1650 rLeuProArgHisLeuAlaProAsnProThrTyrProHisIleTyrProProTyrLeu1 1670
DB 2379 CCTGCCCGACACCTGGCCCGCCACCCACCTACCGGACCTGTACCCACCTACCTCAT 2438
QY 1670 eArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyr1 1690
DB 2439 CCGCGCTACCCCGACACGCGCGGCTGGAGAACCGGACAGACCATCATCAATGACTACAT 2498
QY 1690 eThrSerGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLe 1710
DB 2499 CACCTCGCAGCAGATGTCACACACAGGCCACCGCATGGCCGAGCGAGCTGATGTCT 2558
QY 1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgG1 1730
DB 2559 GAGGGGCTCTCGCCCGGAGTCTCGCTGGCCTCAACTAGCTGCGGGTCCCGAGG 2618
QY 1730 yIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyTh 1750
DB 2619 CATCATGACCTGTCCCAAGTGCACACCTGCTGTGCTCGTCCCGCCGACACAGGCAC 2678
QY 1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSe 1770
DB 2679 CCCAGCACCGCCATGACCGCTTGCCTACCTCCCGACCGCGCCCGAGCTTCAGCAG 2738
QY 1770 rArgHisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrTh 1790
DB 2739 CCGCCACAGAGCTCCCGCTCTCCCGAGGAGTCCAAACACTTGACAAACCAACCAC 2798
QY 1790 rThrSerSerSerGluArgLysArgAspArgGluArgAspArgAspArgGluAr 1810
DB 2799 CAGCTCTCTGTCGAGCGGAGCGAGACCGGATCGAGAGCGGAGCGGATCGGAGCG 2858
QY 1810 gGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleThrArgProG1 1830
DB 2859 GGAAGTCTCATCTCTCAGCTCCACCGACGCGTGGAGACGCGACCATCTGGAGACCTGG 2918
QY 1830 yThrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlySerSerSerAr 1850
DB 2919 TACAGACAGACGAGCGGACGAGCGGACGAGCGGGGGTGGGGGAGCAGCAGCGCG 2978
QY 1850 gProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
DB 2979 CCGCGCTCCCTCCATGCCACAGCATCTGCCCATCTCCCTCGGACCCGAGGATGC 3038
QY 1870 aLeuGlnGluArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVa 1890
DB 3039 CCTCCAGACAGACCCAGTGTCTTCAAAACAGGCATGAGGGTATCATCACCGCTGT 3098
QY 1890 lGluProSerLysProThrValLeuArgSerThrSerSerSerProValArgProAl 1910

Db 5371 AGCTCACCGCTGTCCTCCAGGAGGCCACCTCACCTAGCTAAACCAACCTGCACATCTTCA 5430
Qy SerGluArgGluArgAspArgGluArgAspArgGluArgGluLysSer 1813
Db 5431 TCGAGCGGGAACGGGAACGTGAGCGGGAACGAGAC 5472
Qy IleLeuThrSerThrThrValGluHisAlaProIleTrpArgProGlyThrGluGln 1833
Db 5473 ATCTCTACGTCTACCTACAGTGGAGCANTGCCACCATCTGGAGACCTGTGTAGCGAGCAG 5532
Qy SerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSer 1853
Db 5533 AGCAGCGGGCT 5568
Qy HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGln 1873
Db 5569 CACACC 5622
Qy ArgProSerValLeuHisAsnThrGlyMetLysGlyIleThrAlaValGluProSer 1893
Db 5623 AGGCCCATGTGTGTCACACAGCATGAGGGCGTGTACCTCCGTGGGAACCCCGC 5882
Qy LysProThrValLeuArgSerThrSerThrSerProValArgProAlaAlaThrPhe 1913
Db 5683 ACGCCACCGTCTGAGTCCACCTCCACCTCTTGGCTGTGCGCCAGCTGCCACATTC 5742
Qy ProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMet 1933
Db 5743 CCACCTGCCACCCACTGCCACTTGTGTGGCACCCTTGAAGGGGTCTACCCCTACCTCATG 5802
Qy GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAla 1953
Db 5803 GAGCCGTCTCTGTACCCAGGAGACCTCTCGGGTCGCCGCCGCCGAGCGGCCCGGTG 5862
Qy AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
Db 5863 GACGGTGGCCATGCTCTTCCACCAACCCCGCGCCGG 5913
Qy SerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
Db 5914 TCACCCAGCAGAGCTCCGAGCCCGATCCCTAGCACCCCCCTCCAGCTCCAGCCACAGCC 5973
Qy IleAlaArgThrProAlaLysAsnLeuAlaProHisAlaSerProAspProAla 2013
Db 5974 ATCCCGCGCACCCAGCAAGAGCTTGCACCCACCATGCGCAGTCCGGAACCCCGCGG 6033
Qy ProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIle 2033
Db 6034 ---CCCACCTCGGCTCAGATCTGCACCGAGAAAGACTCAAAGTAAACCTTTTCCATC 6090
Qy GlnGluLeuGluLeuArgSerLeuGlyTyrHis---GlySerSerTyrSerProGluGly 2052
Db 6091 CAGGAATTGGAACCTCCGTTCTCTGGTTTACCACAGTGGAGCTGCTACAGCCCGATGGG 6150
Qy ValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Db 6151 GTGAGCGCCATCAGCCCGGTGAGTCCCCAGCTGACCCACGACAGAGGGCTCTCCAA 6210
Qy HisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092
Db 6211 CCTCTGGAGAGCTAGAGAGAGCCACTTGAAGGGGAGCTGCGGCACAGCCAGCCAGGC 6270
Qy ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer 2112
Db 6271 CCCATGAAGCTCAGCGGGAGGGCTGCCCATCTCCACATCTCGCGGCCACTGCCGAGAGC 6330
Qy GlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
Db 6331 CAGCCCTCATCCAGCCCACTCTCCAGACTGCCCGAGGCATCAAAAGTCCACAGAGGGTG 6390
Qy ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisPro 2152

Db 6391 GTCACCCCTGGCTCAGCACATCATCGAGAGGTCAATTACGAGACTACACCCCGCACCCG 6450
Qy GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPro 2172
Db 6451 CAGACCTAGTGGCCCTTCCCGCCCTCTCTACTCTCTTCCCGGAGCAGCTGCCT 6510
Qy ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAla 2192
Db 6511 GTGCTGGATCTTCGCGCCGCCACCCAGTGCCTCTACTCCACCCCGACCATGGCACC 6570
Qy ProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThr 2212
Db 6571 CCAGCCCGGGATCCCGCCACAGTGAAGGGGGAAGGTCCCCAGAACCCAGCAAAACA 6630
Qy SerValLeuGlyGlyGlyLeuAspGlyIleGluProValSerProProGluGlyMetThr 2232
Db 6631 TCGGTCTTGGGACAGTAGTGGATGCCATTGAGCCCTGTGTCCACCCAGAGGGCATGACT 6690
Qy GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThr 2252
Db 6691 GAGCCAGGACATCTCGGAGCGCTGTGTACCCACTGCTGTATCGAGACGGGAACAGGC 6750
Qy GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
Db 6751 GAGCCC---AGGATGGGCTCTAAGTCTCCAGGCAACACCCAGCGCGCCAGCTTCTTC 6807
Qy SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLys 2292
Db 6808 AGTAAGTGTACTGAGAGAACTCCGCATGTGTGAGTCTGAAGAGCAGGAGATCAACAG 6867
Qy LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
Db 6868 AAATCAACACACCCACCAACCGGAACGAGCCAGATAATAATTTGGCCAGCTTGGACGAA 6927
Qy IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db 6928 ATCTTCAACATGCCCGCATCACTGGAGCAGGCGCTTATGACCTGTAGAAGCCAGCGGTG 6987
Qy GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGly 2352
Db 6988 CAGAACACCCAGCAGCAACATGGGGTAGAGCCATTTATTAGAAGGCACTCATGGGT 7047
Qy LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Db 7048 AAATATGATCAGTGGGAAGAGCCCGCGCTCGGCGCAATGCTTTTAAACCTCTCAAT 7107
Qy AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db 7108 GCCAGCGCCAGTCTGCCCGCTGCTGTATGCCATACCACTGCTGACGAGCGAGTGAC 7167
Qy HisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
Db 7168 CAGCAGCTCACTCCCGAGGTGGAGTGGGAAGCCAAAGGTCTCTGCGACAGCTAGCAGC 7227
Qy ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db 7228 CGAAAGCCCAAGTCGCCAGCACCGAGCTAGCTCGGAGAGCCAGCCCTTCTGTCTCC 7287
Qy SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGlu 2451
Db 7288 TCAGTACACTCAGAGGGGAGCTGCAATCGCCGAACACCACTCAACACCGGTGTGGAG 7347
Qy AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
Db 7348 GACCGGCTCTATCTGAGGGGTCCACGCCATTCCTTACAACTTTGATTATGAGGCTA 7407
Qy GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491
Db 7408 CAGCAGGTGTCTATGGCTCCCGCCCGCCACCTTGGCGGAGCGGCGGCCCTA 7467
Qy AlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLysSerGlnTyrGlu 2511
Db 7468 GCTGGTCCCCACCAACGCTGGAGTGGAGGCCCAAGCCACTGTGTGTTCACAGTATGAG 7527

1058 ProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyr 1077
1078 AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuPro 1097
3343 CCTCCAGGGAGGTGATCAGACTTCCACAGCGCT--GACCT--CTCTTCTCTTAC 3396
3397 ACACCCCGGTCACCGCTGCTCTGGGCTCCACGATAGTCCCGGCGCTCTCTGCCA 3456
1098 ArgProProThrIleSerAsnProProLeuIleSerSerAlaLysHisProSerVal 1117
3457 CGTCCCCC--ATCTAATACCCCCACCCCTCATCTCTCTGCCAAGCATCCCGCGTA 3513
1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyr 1137
3514 CTTGAGAGGCAGCTGGGTGCCATCTCCACAGGGATGTCAGTCAGCTTCGTGTGCCTCAC 3573
1138 SerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
3574 TCAGAGCATGCCAAG--CCATGGGCGCTCTCACCATGGAGCTGCCCTTGGCGTGGAC 3630
1158 ProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGln 1177
3631 CCTAAGAAGCTG----- 3642
1178 AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly 1197
3643 -----GGG 3645
1198 ThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgVal 1217
3646 ACAGCACT--GGCTCCGCCACACAGTGGAGCATCACCAGGGCCT--CCCAGTACCCGGCT 3703
1218 ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
3704 GCAGAGCGCCCC--ACCTACAGAGGCTCTATCACCCACGC--ACGCCCGCAGACGTC 3756
1238 LeuTyrLysGlyThrIleThrArgIleIleGlyLeuAspSerProSerArgLeuAspArg 1257
3757 CTCTACAGGGTACCATCAGCAGATGCTCGGTGAGGACAGCCCAAGTCGCTTGACCGG 3816
1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisVal 1277
3817 GCACGAGAGACACCTCGCCCAAGGGCCATGTCATCTATGAGGCGCAAGAAAGGCCACGTC 3876
1278 LeuSerTyrGluGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
3877 CTATCTTAAAGGTGTATGTCGTCTACAGTGTCTTAAGGAGGATGGAAGGAGCAGC 3936
1298 SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
3937 TCGGGCCACCCATGAGACTGCGGCCCTTAACGCACCTATGACATGATGGAGGCGCT 3996
1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
3997 GTAGGCGAGACTGTCACTCAGCAGCATAGAGGACTCATGCGCGCGCCATC--CCT 4053
1338 GluArgHisSerProHisHisLeuLysGluGlnHisIleArgGlySerIleThrGln 1357
4054 GAGCAGCACAGCCCC--CACCTCAGGAGCAGCATCACATCCGAGGCTCCATCAGCGAA 4110
1358 GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeu 1377
4111 GGCATCCCGAGGTCTATGTGGAGGCGCAGGAGACTTATCGCGGGAGGCCAAGCTC 4170
1378 LeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyr 1397
4171 TTGAAGCGAGAGGAGCACACCTCCGCCACACCCACCTCGGAGCCTGACTGAGACCTAC 4230
1398 LysThrGln-----AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeu 1414
4231 AAGCCCCGGCCCTGGACCTCTGGGTCCCCCTGAAGCTGAAGCCGACTCACGAGGGGTGTG 4290

1415 ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
4291 GTAGCAACTGTGAAGGAGCGGCGCTCTATCATGAGATCCCGAGAGAGGAGTGGCG 4350
1435 HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly 1454
4351 CGCACACCTGAGCTACCTCGCCACACCGCTCTGAAGGAGGGTTCATACCCAGGGC 4410
1455 ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArg 1474
4411 ACCCACTCAGTACGACTCTGGGGCACCTCCAGTGGACCAAGAAACACGACGTGGC 4470
1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
4471 TCCATCATCGGAGCCCGCGCGCTTTCCTCGCTCCCTGCACCGCTGGACATAATGGCT 4530
1495 AspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLysSerArgProGlyThr 1514
4531 GACGCCCGGGCACTGGAGCGTGCCTGCTATGAAGAGAGTCTGAAGAGCCGGTCAGGACC 4590
1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
4591 AGCAGTGTGTCAGGGGCTCCATCACACGTCGGGCTCCAGTCTGCTGCTGAACTGGGC 4650
1535 LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeu 1554
4651 AAGCCACGCAAGCCCACTGACTTACGAAGACACACGGGCACTTCCACAGTCACTTG 4710
1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluLysSerLeu 1574
4711 CCACGTGGTCTCCCTGTGACACGAGGAGCCACCGCTCAGGAAAGCAGCGCTC 4770
1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
4771 CTATCCAGCAGAGCGTCCACGAGCCGAAAGTGCATCTACACCCCGGAGATCGCCAG 4830
1595 SerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeu 1614
4831 TCCCCACACAGCACTGTGCGCAGCAGCACCCCTCACCCCATCTCCCTTATGAGCACTTG 4890
1615 LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro 1634
4891 CTCGGGGGCTGACTGGTGTGACCTGTACCGTGTGCATCCCATTTGGCCTTTGACCCC 4950
1635 ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaIaTyrTyrLeuProArg 1653
4951 ACCTCATACCCGAGGGATCCCTCTGGAAGCAGCAGCGACCTACTACTCTGCCCGG 5010
1654 HisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyr 1673
5011 CACTTGGCCCCCAGCCCCACCTACCCACACCTGTACCCACCTTACCTCATCCGCGGTAC 5070
1674 ProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGln 1693
5071 CCTGACACCGCGCCCTGGAGAACCCCGCAGACCATCATCAATGACTACATCATCGCAG 5130
1694 GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
5131 CAGATGCACACACGCTGCTCCGCGCATGGCCAGCGTGTGCATGCTGAGGGGTCTG 5190
1714 SerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp 1733
5191 TCACCGCAGAGTCTCTCGCTGGCCCTCAATTATTCCGCTGGCCCAAGAGGACATTATCGAC 5250
1734 LeuSerGlnValProHisLeuProValLeuValProProThrProThrProAlaThr 1753
5251 CTGTCCCAAGTGCACACCTGCGCGTGGTGGTCCCAACAGCAGGACACCCCTGCGACC 5310
1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
5311 GCATCGACGCGCTTGCCTACCTCCCACTGCGCCCGCCACCTTTCAGCAGCGCCACAGT 5370
1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSer 1793

| | | | | |
|----|------|---|------|--|
| Db | 1180 | ATCCGCAAGCAGCGGAGCTGCAAGGAGCGCATGACAGAGGCTGGCGCAGCGTGCAGT | 1239 | |
| Qy | 360 | GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu | 379 | |
| Db | 1240 | GGGCTCTCCATGTCGGCTGCCGAGTACAGATGAGGTTCTGAGATCATTTGATGCTTG | 1299 | |
| Qy | 380 | SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeu | 399 | |
| Db | 1300 | TCTGAGCAGGAGAACCTGGAGAACGATGCGCAGCTGGCCGCTGATCC--GCATGTTG | 1357 | |
| Qy | 400 | TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro | 419 | |
| Db | 1358 | TACGACGC-GACCAGCAGAGATCAAGTTTCAACAATGATGACTCATGGATGACCCC | 1416 | |
| Qy | 420 | MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr | 439 | |
| Db | 1417 | ATGAAAGTCTACAGGACCGCTCAGTTTACCACATGTGGAGCAGGAGGAGGCACCC | 1476 | |
| Qy | 440 | PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu | 459 | |
| Db | 1477 | TTCCGTGAGAAGTTTATGACGACCCCTAAGAACCTTTGGCCTGATTGCTCATTCCTGGAG | 1536 | |
| Qy | 460 | ArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyr | 479 | |
| Db | 1537 | AGAAAGCGGTGCTGAGTGTCTCTATTACTTACTCTGACCAAGAGATGAATAATTAC | 1596 | |
| Qy | 480 | LysSerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGln | 499 | |
| Db | 1597 | AAGAGCTTGTGAGCGGAGCTATCGCGCGCTGCGCAAGAGCCAGCAGCAGCAGCAG | 1656 | |
| Qy | 500 | GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGlu | 519 | |
| Db | 1657 | CAACAACAGCAGCAGCAGCAGCAG-----ATGGCAGCGAGCAGCGCAGGAG | 1704 | |
| Qy | 520 | LysAspGluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGlu | 539 | |
| Db | 1705 | AAGAGGAGAGGAGGAGGAGGAGGAGCGCCAGAGGAGGAAGAGCAGGATGCGGAG | 1764 | |
| Qy | 540 | AsnAspLysGluAspLeuLysGluLysThrAspAspThrSerGlyGluAspAsp | 559 | |
| Db | 1765 | AACGAGAAGGAAGAACTCAGCAAGAGAGAGCAGACGACACTTCTGGCGAGGACAACGAT | 1824 | |
| Qy | 560 | GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgLys | 579 | |
| Db | 1825 | GAGAAAGAGCGCTGGCTTCCAAAGGCCGCAAACTGCCAACACAGCCGCGCGCAA | 1884 | |
| Qy | 580 | GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln | 599 | |
| Db | 1885 | GGCCGTATCCGCGCTCCATGGCCAAACGAGGCCAACCATGAGGAGACAGCCACCCACAG | 1944 | |
| Qy | 600 | GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTTrpThrGluGluGlu | 619 | |
| Db | 1945 | CAAAATTCAGAGCTGGCTTCCATGAGATGAACAGAGATTTCTCGCTGGACATGAGGAAGAG | 2004 | |
| Qy | 620 | MetGluThrAlaLysLysGlyLeuGluHisGlyArgAsnTrpSerAlaIleAlaArg | 639 | |
| Db | 2005 | ATGGAGACAGCAAGAGGCTCTCTGGAACATGGGAGAACTGGTCAGCCATTGCCCGC | 2064 | |
| Qy | 640 | MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg | 659 | |
| Db | 2065 | ATGTTGGGCTCCAGACCGCTGCCAGATGTAGAACTTCTACTTCAACTACAAGAAGAGG | 2124 | |
| Qy | 660 | GlnAsnLeuAspGluIleGlnGlnHisLysLeuLysMetGluLysGluArgAsnAla | 679 | |
| Db | 2125 | CAGAACTTGGAGAAATCTTTCAGCAGCACAAGCTTAAGATGCAAGAGGAGGAGNACGCT | 2184 | |
| Qy | 680 | ArgArgLysLysLysAlaProAlaAlaSerGluGluAlaAlaPheProProVal | 699 | |
| Db | 2185 | CGGAGGAG | 2244 | |
| Qy | 700 | ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGlu | 719 | |
| Db | | | | |

| | | | | |
|----|------|--|------|--|
| Db | 2245 | GCTGAGGACGAGAGATGGAAGCATCAGCGCGCAAGTGCCAATGAGAAAGAGCTGCGGAG | 2304 | |
| Qy | 720 | GluAlaGluAlaLeuHiAlaSerGlyAsnGluValProArg--GlyGluCysSerGly | 738 | |
| Db | 2305 | GAGGAGAGAGCTCTCAGGCCTCTCGGAATGAGTTCCAGAGTTGGGAGTGCAGTGC | 2364 | |
| Qy | 739 | ProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAla | 758 | |
| Db | 2365 | CCAGCTGTGTGTCACACACAGCTCTGACTGAGAGTGTCCCATCCCGCGTTTCAGAGCC | 2424 | |
| Qy | 759 | AlaLysAspThrGlyGlnAsnGlyProLysProProAlaThr-----LeuGlyAlaAsp | 776 | |
| Db | 2425 | ACGAAAGCACT-----GGGCTAAACCACTGGCAGCTGAAGCATTTGCCCTGCC | 2475 | |
| Qy | 777 | GlyProProGlyProProThrProProArgThrSerArgAlaProIleGluPro | 796 | |
| Db | 2476 | ACCGAGCCACCTTTCTCT-----CTCCAGAAACCGGACGAGCCCTCTGCTGAGCCC | 2529 | |
| Qy | 797 | ThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSer | 816 | |
| Db | 2530 | TCCCCAGTCCCTGATGCGCAGTGGCCACCATCCCGAGAGCTTCC--CCATCACCTGCC | 2586 | |
| Qy | 817 | AlaProProValValProLysGluGluLysGluGluThrAlaAlaAlaProPro | 836 | |
| Db | 2587 | GCACCCCGCGTACTGTGACCAAGGATGAACAGAGAGCCCGGCTCTCCAGCTCCCCAG | 2646 | |
| Qy | 837 | ValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGly | 856 | |
| Db | 2647 | ACGAAAGTCCAGAGGAGGAGAGTCTGAGGCCGAGGAG-----ATCGATGTGGGA | 2697 | |
| Qy | 857 | LysAlaGluGluPro-----VallysSerGluCysThr | 867 | |
| Db | 2698 | AGCCAGGAGGAGCGCGAGGCTCTGAGGAGCGCCCGGAGAGTGTAAAGAGTGACCAAG | 2757 | |
| Qy | 868 | GluGluAlaGluGluGlyProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThr | 886 | |
| Db | 2758 | GAGGAGCCGAGAGAGAGGCTTGAAGCAAGCCAGGCGCAGAGGCCATTGAACTGTG | 2817 | |
| Qy | 887 | AlaGluGlyAlaLeuLysAlaGluLysLysGluGlyLysGlyArgAlaThr---905 | 905 | |
| Db | 2818 | TCTGAGGACCACTTAAAGTGGAG-----GAGCTGGTAGCAAGCAGCTGTGACCAAG | 2871 | |
| Qy | 906 | AlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGlu | 925 | |
| Db | 2872 | GGTTCCAGCTCAGTGCACCCAGGACAGTCACTCCAGTGCCACCTGCAGTCCCGATGAG | 2931 | |
| Qy | 926 | ValAspGluAlaGluGlyGlyAspLysAsnArgLeuSerProArgProSerLeuLeu | 945 | |
| Db | 2932 | GTGACCAACCCAGAGAGGTGACAGGCGAGGCTGTGTCCACCAAGGCCAGCCTCTC | 2991 | |
| Qy | 946 | ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLysGln | 965 | |
| Db | 2992 | ACCCCGCTGGAGATCCCGGCGCAGTACCTCCGCCCCAGAAAGCGCTGGACCTGAAGCAG | 3051 | |
| Qy | 966 | LeuLysGlnArgAlaAlaIleProProIleGlnValThrLysValHisGluProPro | 985 | |
| Db | 3052 | CTGAAGCAGGAGCAGCGCCCTATCCCTTATC---GTCACCAAGGTTCATGAGCCCCC | 3108 | |
| Qy | 986 | ArgGluAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeu | 1005 | |
| Db | 3109 | CGGAGGACACAGTACCCCCCAAGCCAGTTCCTCCCTGTCTCCACCCAGCAGCAGCTA | 3168 | |
| Qy | 1006 | GlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSer | 1025 | |
| Db | 3169 | CAGCAGAGGAGTACGTGTCTCAGCAGCTCGGAGGAAGTCCAGTGGCAAGTCCCGCAGC | 3228 | |
| Qy | 1026 | ProAlaProProAlaAspLysGlu-----AlaPheAlaAla | 1037 | |
| Db | 3229 | CCAGTGTCTCTCGGAGAAAGGAGGAGAGAAACCCGATCTTTCGGCTTTTCCCAACT | 3288 | |
| Qy | 1038 | GluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVal | 1057 | |
| Db | 3289 | GAGGCGCA-AAGTACCGACTGAGCCCCCAGCTGTGTCTCATCGCT---GCTTCCCAT-- | 3342 | |


```
Db 7249 CAGCGAGGTGTATGTCCTCCCGCCCGCCACCTCGGCTTGGCGAGGACGCGGCCCTTA 7308
QY 2492 AlaGlyProHisHisAlaTTPAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
Db 7309 GCTGGTCCCCACACCGCTGGATGAGAGGCCCAAGCCACTGCTGTGTTCACAGTATGAG 7368
QY 2512 ThrLeuSerAspSerGlu 2517
Db 7369 ACACTCTCGGACGCGAG 7386

RESULT 5
AC62451
ID ACA62451 standard; cDNA; 8544 BP.
XX AC
XX ACA62451;
XX
DT 12-AUG-2003 (first entry)
DE cDNA encoding mouse nuclear receptor corepressor SMRTE.
XX
KW Mouse; es; gene; SMRTE; nuclear receptor corepressor; gene therapy;
KW tissue typing; cancer.
XX
OS Mus musculus.
XX
PH Key Location/Qualifiers
FT CDS 160..7548
FT /*tag= a
FT /product= "SMRTE"
XX
XX US2003027137-A1.
XX
XX
PD 06-FEB-2003.
XX
XX 27-MAR-2001; 2001US-00819104.
XX
XX 29-MAR-2000; 2000US-0193138P.
XX
XX (CHEN/) CHEN J D.
XX
XX Chen JD;
XX
XX WPI; 2003-466139/44.
XX
XX P-PSDB; ABU61813.
XX
XX
XX New SMRTE proteins and nucleic acids, useful in gene therapy, predictive
XX medicine, therapeutic or prophylactic treatment, chromosome mapping,
XX tissue typing and in forensic biology.
XX
XX Claim 2; Page 56-65; 90pp; English.
XX
XX The invention relates to an isolated SMRTE nucleic acid molecule. The
XX nucleic acids are useful in gene therapy, as hybridisation probes for
XX identifying SMRTE-encoding nucleic acid molecules and as primers for
XX amplifying of SMRTE nucleic acid molecules. The polypeptides are useful
XX as immunogens to raise anti-SMRTE antibodies. The SMRTE molecules are
XX useful as targets for discovering and developing modulating agents to
XX regulate a variety of cellular processes, in screening assays, in
XX predictive medicine, in therapeutic or prophylactic treatment, in
XX chromosome mapping, tissue typing and in forensic identification of a
XX biological sample. Modulators of SMRTE are useful for treating or
XX preventing a condition associated with aberrant SMRTE protein or nucleic
XX acid expression or activity, such as cancer. The present sequence
XX represents cDNA encoding the mouse nuclear receptor corepressor SMRTE
XX
XX Sequence 8544 BP; 1970 A; 2882 C; 2378 G; 1314 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0 Length: 8544
Score: 10832.50 Matches: 2149
Percent Similarity: 87.94% Conservative: 90
Best Local Similarity: 84.41% Mismatches: 197
Query Match: 81.97% Indels: 117
```

```
DB: 7 Gaps: 33
US-09-522-753-5 (1-2517) x ACA62451 (1-8544)
QY 1 MetSerGlySerThrClnLeuValAlaGlnThrTTPArgAlaThrGluProArgTyrPro 20
Db 160 ATGTCTAGGATCCACACAGCCTGTGGCACAGACATGCGGGCTGCTGAGCCCGCTACCCA 219
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 220 CCCATGGCATCTCTACCGGTGAGATAGCCGGTCCACACGAGCGTGGGGCTGCTT 279
QY 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 280 GAGTACCAACACCCCGCTGACTACACTCACACCTGTCAACCCGTTCCATCATCCAG 339
QY 61 ProGlnArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 340 CCACAGAGAGCGGCCCTCACTGTCTCAGAGTTCCAGCCTGGGAGTGAACGGTCTCAG 399
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
Db 400 GAGCTCCACTGCGCCCTGAGTCCCGCACGTTCTCTGCTGAGTGGGCAAGCCCGACATA 459
QY 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 460 GAATTCACCGAGAGCAAGCGCCCGCTGGAGCTACTACCCGATACCTGTCTGCGCCCA 519
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 520 TCACCCCTGCTGGCCACTGGCGAGCGAGTGGGTCTTGAAGACCTTACCAAGGACCGTAGC 579
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeu 160
Db 580 CTGGCAGGCAAGCTGGAGCCTGTGTCACTCCCGTCCCGCCGACGCTGAGCTT 639
QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsn--MetAspArgVal 179
Db 640 GAGCTGGCGCCATCTCGACTGTCCAGAGGAGGAGCTGATCCAGAACAGATTGGACCGGTG 699
QY 180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGln 199
Db 700 GACCGTGAGATCACCATGGTAGAGCAGCAGATCTCCAAGCTGAAGAGAGCAGCAACAG 759
QY 200 LeuGluGluAlaAlaLysProGluProGluProGluProValSerProProProIle 219
Db 760 TTGGAGGAGAGCGCCCAAGCCCGCCAGAACCCGAGAGCCTGTGTGCCACCCACCCATA 819
QY 220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu 239
Db 820 GAATCAAGCACCGAAGCCTGGTCCAGATCATCTACGATGAGAACCGGAAGAACCGGAA 879
QY 240 AlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGln 259
Db 880 GC CGCACACCGGATCTCTAGAACGCTGGGGCCCGCAGGTGGAGCTGCTCTGTACAACAC 939
QY 260 ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLys 279
Db 940 CCCTCTGACACACCGCAGTACCATGAAACATCAAAATAAACCCAGCGGATGCGGAAGAAG 999
QY 280 LeuIleLeuTyrPhenylsArgArgAsnHisAlaAArgLysGlnTyrLysGlnLysPheCys 299
Db 1000 CTGATCTTGTACTTTAAGCGGAGAACCCAGCGCGCAGCAGTGGGACACAGCGCTTCTGC 1059
QY 300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsn 319
Db 1060 CAGCGCTATGACCCAGCTCATGGAGCGGTGGGAGAGAGAGGTAGAGCGCATAGAACCAAT 1119
QY 320 ProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db 1120 CC CGCAAGGAGGCGCCCAAGGAGAGCAAGGTGAGGAGTACTTACGAGAAACAGTTCCCGGAG 1179
QY 340 IleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer 359
```


Db 3070 CCAGTGCCTCTCGCCGAGAAAGAGGAGAGAAACCCGCAATCTTTCCGCGCTTTCCCAACT 3129
Qy 1038 GluAlaGlnLysLeuProGlyAspProCysTyrThrSerGlyLeuProPheProVal 1057
Db 3130 GAGGGCCA-AGCTACCGACTGAGCCCAACCGTGTCTATCGCT---GCCTTCCCAT-- 3183
Qy 1058 ProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyr 1077
Db 3184 CTTCCACGGAGGTGATCAAGACTTCCACACGGCT---GACCT---CTCTTCTCTCTAC 3237
Qy 1078 AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuPro 1097
Db 3238 ACACCCCCCGGTACCCGCTGCTCTGGGCTCCACGATAGTCCCGGCGGCTGCTGCCA 3297
Qy 1098 ArgProProThrIleSerAnProProProLeuIleSerSerAlaLysHisProSerVal 1117
Db 3298 GGTCCCCC---ATCTCTAACCCCCCAACCCCTCATCTCTCTGCCAAGCATCTCCGGCGTA 3354
Qy 1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyr 1137
Db 3355 CTTGAGAGCAGCTGGGTGCATCTCCAGGGGATGTCACTGCAGCTTGTGTGCTCTAC 3414
Qy 1138 SerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
Db 3415 TCAGAGCATGCCAAG---CCCATGGCCCTCTCACCATGAGCTGCCCTTGGCGTGAGC 3471
Qy 1158 ProLysLeuAlaProPheSerGlyValLysGlnLeuSerProArgGlyGln 1177
Db 3472 CCTAAGAGCTG----- 3483
Qy 1178 AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly 1197
Db 3484 -----GGG 3486
Qy 1198 ThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgVal 1217
Db 3487 ACAGCACT-GGCTCCGCCACACAGTGAAGCATCAACCAAGGGCT-CCCAGTACCCGGCT 3544
Qy 1218 ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
Db 3545 GCAGAGGGCCCC-----ACGTACAGAGGCTCTATCACCCACGC-ACGCCCGCAGAGCTC 3597
Qy 1238 LeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArg 1257
Db 3598 CTCTACAAGGTACCATCAGCAGATCGTGGTGAGGACAGCCCAAGTGCCTTGACCGG 3657
Qy 1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGlnLysLysGlyHisVal 1277
Db 3658 GCACGAGAGACACCTCGCCCAAGGGCCATGTCTATGAGGGCAAGAAAGGCCACGTC 3717
Qy 1278 LeuSerTyrGlnGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
Db 3718 CTATCTTATAGGTGTATGTCCGTGTCAAGTGTCTTAAGAGGATGAAGAGCAGC 3777
Qy 1298 SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetGluGlyArg 1317
Db 3778 TCGGGCCCAACCCATGAGCTGCGGCCCTTAACGCACCTATGACATGATGAGGGCCGT 3837
Qy 1318 ValGlyArgAlaIleSerSerAlaSerIleGlyLeuMetGlyArgAlaIleProPro 1337
Db 3838 GTAGGCAGGACTGTCACTCAGCAGCATAGAGGACTATGAGGGCGCGCATC---CCT 3894
Qy 1338 GluArgHisSerProHisHisLeuLysGlnHisIleArgGlySerIleThrGln 1357
Db 3895 GAGCAGCACAGCCCC---CACTTCAAGAGCAGCATCATCTCGAGGCTCCATCAGCCAA 3951
Qy 1358 GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeu 1377
Db 3952 GGCATCCCGAGGTCTATGTGAGGGCGCAGGAGCTACTTACGGGGGAGGCCAAGCTC 4011
Qy 1378 LeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyr 1397
Db 4012 TTGAAGCGAGAAGGACACACCTCTCCCCACCCACCTCTCGGACCTGAGCTGAGACCTAC 4071

Qy 1398 LysThrGln-----AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeu 1414
Db 4072 AAGCCCGCGCCCTCGGACCTCTGGTCCCTGAAAGCTGAAGCCGACTCACGAGGGTGTG 4131
Qy 1415 ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
Db 4132 GTAGCAACTGTGAAGAGCGCGCGCTCTATTCATAGATCCGAGAGAGGAGCTGCGC 4191
Qy 1435 HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly 1454
Db 4192 CGCACACCTGAGCTACCCCTGSCACACCGCTCTGGAAGAGGGTTCATCACCCAGGGC 4251
Qy 1455 ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArg 1474
Db 4252 ACCCCACTCAAGTACGACTCTGGGCGACCTCCACTGGCACCAAGAAACACGAGCTGCGC 4311
Qy 1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db 4312 TCATCATCGGAGCGCGCGCGCTTTCCTGCCCTGCACCGCTGGACATAATGGCT 4371
Qy 1495 AspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThr 1514
Db 4372 GACCCCGGGCACTGGAGGCTGCTGTATGAAGAGAGTCTGAAGAGCGGTGAGGAGC 4431
Qy 1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
Db 4432 AGCAGTGTGTGAGGCGCTCCATCACAGCTGGGGCTCCAGTCTGCTGCTGAACCTGGGC 4491
Qy 1535 LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeu 1554
Db 4492 AAGCCACGCAAGCCCACTGACTTACGAAGACCCACGGGGCACCTTCCACAGCTCACTG 4551
Qy 1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
Db 4552 CCAGTGTGCTCCCTGTGACAGGAGGAGCCACCGCCGCTTTCAGNAGGAGCGCTC 4611
Qy 1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
Db 4612 CTATCCAGCAAGCGCTGCCAGGACCGGAAGTGCATCTACACCCCGGAGATCGCCAG 4671
Qy 1595 SerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeu 1614
Db 4672 TCCCCACAGACACTGTGCCCGAGCACACCTCCACCCCTATCCCTATGAGCACTTG 4731
Qy 1615 LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro 1634
Db 4732 CTCGGGGGTGACTGTGTGACCTGTACCTGTGTCATCCATCCATTTGGCTTTGACCCC 4791
Qy 1635 ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaIleTyrTyrLeuProArg 1653
Db 4792 ACCTCATACCCCGAGGGATCCCTCTGGAAGACGAGCAGCGCCCTACTACCTGCCCGG 4851
Qy 1654 HisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyr 1673
Db 4852 CACTTGGCCCCCAGCGCCCACTACCCACACCTGTATCCCACTTACCTCATCCCGGGCTAC 4911
Qy 1674 ProAspThrAlaAlaLeuGluAsnArgGlnThrIleLeuAsnAspTyrIleThrSerGln 1693
Db 4912 CTTGACACGGCGGCTTGGAGAACCCCGCAGACCATCATCATGACTACATCATCTCGCAG 4971
Qy 1694 GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db 4972 CAGATGACACCAACCGCTCCGCGCATGGCCCGAGCGTGTGCATGCTGAGGGGTCTG 5031
Qy 1714 SerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp 1733
Db 5032 TCACCGCAGAGTCTCTCGCTGGCCCTCAATTAATTCGCTGGCCCAAGAGCATTTATCGAC 5091
Qy 1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Db 5092 CTGTCCCAAGTGCACACCTTGGCCCGCTGTGTGTCGCCAACGACGAGCACCTCTGCCACC 5151

Qy 320 ProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db 961 CCGGAAAGGAGGCGCAAGAGAGAGCAAGGTGAGGAGTACTACGAGAAACAGTTCCTCCGAG 1020
Qy 340 IleArgLysGlnArgGluLeuGlnArgMetGlnSerArgValGlyGlnArgGlySer 359
Db 1021 ATCCGCAAGCAGCGGAGCTCGAGAGCGCATGCGAGCAGGGTGGCCAGCGTGGCAGT 1080
Qy 360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluLeuIleLeuAspGlyLeu 379
Db 1081 GGGCTCTCCATGTCGGCTGCCCGCAGTGCAGATGAGGTTCCTGAGATCAATGATGGCTTG 1140
Qy 380 SerGluGlnGlnAenLeuLysGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Db 1141 TCTGAGCAGGAGAACCTGGAGAACGATGCGCAGCTGGCGGTGATCC--GCCATGTTG 1198
Qy 400 TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db 1199 TACGACGC-GACCAGCAGAGGATCAAGTTCATCAACATGATGAGTGCATGATGACCCC 1257
Qy 420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
Db 1258 ATGAGGTCTACAGGACCGTCACTTACCACATGTGGAGCGAGCAGGAGGACACC 1317
Qy 440 PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu 459
Db 1318 TTCGTGAGAAGTTTATGCAGCACCTTAAGAACCTTTGGCCTGATTGCTCATTCCTGGAG 1377
Qy 460 ArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyr 479
Db 1378 AGAAAGACGGTCGTGAGTGTCTCTATTACTTACCTGACCAAGAAAGATGAAAAATTAC 1437
Qy 480 LysSerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGln 499
Db 1438 AAGAGCTTGTGAGCGGAGCTATCGCGCGCTGGCAAGAGCCAGCAGCAGCAGCAG 1497
Qy 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGlu 519
Db 1498 CAACACAGCAGCAGCAGCAGCAG-----ATGCGCAGCAGCAGCAGCAGGAG 1545
Qy 520 LysAspGluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGlu 539
Db 1546 AAGAGGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1605
Qy 540 AsnAspLysGluAspLeuLysGluLysGluLysThrAspAspThrSerGlyGluAspAsnAsp 559
Db 1606 AACGAGAGAGAGAGACTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1665
Qy 560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
Db 1666 GAGAAAGAGCGGTGGCTCCAAAGGCGCAAAACTGCCAACAGCCCAAGGCGCGCAAA 1725
Qy 580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
Db 1726 GGCCTGATACCGCGCTCCATGGCCCAACGAGCGCCAAACCATGAGGAGAGAGAGAGAG 1785
Qy 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGlu 619
Db 1786 CAAAGTTCAGAGCTGGCTTCATGGAGATGAACGAGAGTTCCTCGCTGCATGAGGAGAG 1845
Qy 620 MetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArg 639
Db 1846 ATGGAGACAGCAAAAGAGCCCTCTGGAACATGGGAGGAATGGTTCAGCCATTTGCCCGC 1905
Qy 640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
Db 1906 ATGGTGGGCTCCAGACCGGTGCCAGTGTAAAGAACTTCTACTTCAACTACAGAGAGAG 1965
Qy 660 GlnAsnLeuAspGluIleLeuGlnGlnHisLysLysLysMetGluLysGluArgAsnAla 679
Db 1966 CAGAACTTGACGCAAACTCTTCAGCAGCACAAGCTAAAGATGAGAGAGAGAGAGAGCT 2025
Qy 680 ArgArgLysLysLysAlaProAlaAlaAlaSerGluAlaAlaPheProVal 699

Db 2026 CCGAGGAAGAAGAAGAAGACCCAGCTCGCGCAGCAGGAGACAGCGCTTCCACCTGCC 2085
Qy 700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGlu 719
Db 2086 GCTGAGGACCAAGAGATGGAAGCATCAGGGCCAAATGAGGATGCCAATGAGGAGAGCTGGCGAG 2145
Qy 720 GluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArg---GlyGluCysSerGly 738
Db 2146 GAGGACAGAGCCCTCACAGGCTCTGGGANTGAGGTTCCAGAGTTGGGAGTGAGTGCC 2205
Qy 739 ProIaThrValAsnAenSerSerAspThrGluSerIleProSerProHisThrGluAla 758
Db 2206 CCAGCTGCTGTCACAAACAGCTCTGTACTGAGAGTGTCCCATCCCGCGTTTCAAGAGCC 2265
Qy 759 AlalysAspThrGlyGlnAsnGlyProLysProProIaThr-----LeuGlyAlaAsp 776
Db 2266 ACGAAGGACACT-----GGCCCTAAACCCCACTGGCACTGAAGCATTTGCCGCTGCC 2316
Qy 777 GlyProProGlyProProThrProArgArgThrSerArgAlaProIleGluPro 796
Db 2317 ACCAGCCACTGTTCT-----CTCCAGAGAACCGGAGCAGGAGCCCTGCTGAGGCC 2370
Qy 797 ThrProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSerProSer 816
Db 2371 TCCCGACTCCTGATGCGAGTGGCCACCATCCCGCAGAGCTTCC--CCATCACCTGCC 2427
Qy 817 AlaProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProPro 836
Db 2428 GCACCCCGGCTACTGTGGCAAGAGTGAACAGAGCCCGGCTGTCTCCAGCTCCCGAG 2487
Qy 837 ValGluGluGlyGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGly 856
Db 2488 ACAGAAGATCCCAAGGAGCAGAGTCTGAGCCCGAGAG-----ATCGATGTGGGA 2538
Qy 857 LysAlaGluGluPro-----ValLysSerGluCysThr 867
Db 2539 AAGCAGAGGAGCGCAGGCTCTGAGGAGCGCCCGGAGAGTGTAAAGAGTGAACCAAG 2598
Qy 868 GluGluAlaGluGluGlyProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThr 886
Db 2599 GAGGAGACCCAGAGAGGCTGGAAGACAAAGCCAGGCGCACAGAGGCCATTTGAACCTGTG 2658
Qy 887 AlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr--- 905
Db 2659 TCTGAGGCACCATTAAGTGGAG-----GAGCTGTGAGCAGGAGCTGTGACCAAG 2712
Qy 906 AlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGlu 925
Db 2713 GGTTCAGCTCAGTGCACCCAGGACAGTACTCCAGTGCCACCTGCAGTGCCGATGAG 2772
Qy 926 ValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeu 945
Db 2773 GTGGACGAAACCCAGAGAGGTGCAAGGGCAGGCTGCTGTCAACAGGCCAGCTCTC 2832
Qy 946 ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGln 965
Db 2833 ACCCGGCTGGAGATCCCCGGGCGCAGTACCTCGCCCCAGAGCGCTGGACCTGAGAGCAG 2892
Qy 966 LeuLysGlnArgAlaAlaIleProProIleGlnValThrLysValHisGluProPro 985
Db 2893 CTGAAGCAGCAGCAGCGCCCATCCCGCTATC---GTCACCAAGGTCCATGAGCCCCC 2949
Qy 986 ArgGluAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeu 1005
Db 2950 CGGGAGGACACAGTACCCCAAGCCAGTTCCCGCTGTGCTCCACCCAGCAGACCTA 3009
Qy 1006 GlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSer 1025
Db 3010 CAGCCAGAGGTCAGCTGTCTGAGCAGTCGGGAGGAGTCCAGTGGCAAGTCCCGCAGC 3069
Qy 1026 ProAlaProProAlaAspLysGlu-----AlaPheAlaAla 1037

QY 2472 nAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7542 GCGGGTGTTCATGGCTTCCACCCACCGGCTCCCGGGGAGCGGGCCCTCGC 7601
QY 2492 aGlyProHisHisAlaTTPAspGluGluProLysProLeuLeuCysSerClnTyfGluTh 2512
Db 7602 TGGCCCCCACCAGCCTGGAGAGAGGCCCAAGCCACTGCTGTGTCGAGTACGAGAC 7661

QY 2512 rLeuSerAspSerGlu 2517

Db 7662 ACTCTCCGACGCGAG 7677

RESULT 4

ACA62452

ID ACA62452 standard; cDNA; 7386 BP.

XX ACA62452;

XX 12-AUG-2003 (first entry)

DE Mouse nuclear receptor corepressor SMRTE coding region cDNA.

XX Mouse; ss; gene; SMRTE; nuclear receptor corepressor; gene therapy;

KW tissue typing; cancer.

XX Mus musculus.

FH Key Location/Qualifiers

FT CDS 1..7386

FT /tag= a

FT /partial

FT /product= "SMRTE"

FT /note= "No stop codon given"

PN US2003027137-A1.

XX 06-FEB-2003.

XX 27-MAR-2001; 2001US-00819104.

XX 29-MAR-2000; 2000US-0193138P.

XX (CHEN/) CHEN J D.

XX Chen JD;

XX WPI; 2003-466139/44.

DR P-PSDB; AB061813.

XX New SMRTE proteins and nucleic acids, useful in gene therapy, predictive medicine, therapeutic or prophylactic treatment, chromosome mapping, tissue typing and in forensic biology.

PS Claim 2; Page 71-79; 90pp; English.

XX The invention relates to an isolated SMRTE nucleic acid molecule. The nucleic acids are useful in gene therapy, as hybridisation probes for identifying SMRTE-encoding nucleic acid molecules and as primers for amplifying of SMRTE nucleic acid molecules. The polypeptides are useful as immunogens to raise anti-SMRTE antibodies. The SMRTE molecules are useful as targets for discovering and developing modulating agents to regulate a variety of cellular processes, in screening assays, in predictive medicine, in therapeutic or prophylactic treatment, in chromosome mapping, tissue typing and in forensic identification of a biological sample. Modulators of SMRTE are useful for treating or preventing a condition associated with aberrant SMRTE protein or nucleic acid expression or activity, such as cancer. The present sequence represents the mouse nuclear receptor corepressor SMRTE coding region cDNA

XX Sequence 7386 BP; 1735 A; 2520 C; 2046 G; 1085 T; 0 U; 0 Other;

XX Alignment Scores:

| | | | |
|--|----------|---|------|
| Pred. No.: | 0 | Length: | 7386 |
| Score: | 10832.50 | Matches: | 2149 |
| Percent Similarity: | 87.94% | Conservative: | 90 |
| Best Local Similarity: | 84.41% | Mismatches: | 197 |
| Query Match: | 81.97% | Indels: | 117 |
| DB: | 7 | Gaps: | 33 |
| US-09-522-753-5 (1-2517) x ACA62452 (1-7386) | | | |
| QY | 1 | MetSerGlySerThrGlnLeuValAlaGlnThrTTPArgAlaThrGluProArgTyrPro | 20 |
| Db | 1 | ATGTCTAGGATCCACACAGCCTGTGGCACACAGATGGCGGGCTGCTGAGCCCCGCTACCCA | 60 |
| QY | 21 | ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu | 40 |
| Db | 61 | CCCATGGCATCTCTCTACCCGGTGCAGATAGCCCGGTCCACACGACGCTGGGGCTGCTT | 120 |
| QY | 41 | GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln | 60 |
| Db | 121 | GAGTACCAACACCCCGCTGACTACACCTCACACCTGTCAACCCGGTTCCATCATCCAG | 180 |
| QY | 61 | ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln | 80 |
| Db | 181 | CCACAGAGAGGGGGCCCTCACTGTCTCAGAGTTCAGCCTGGGAGTGNACGGTCTCAG | 240 |
| QY | 81 | GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet | 100 |
| Db | 241 | GAGTCCACCTGCGCCCTGAGTCCCGCACGCTTCTGCTGAGCTGGGCAAGCCCGACATA | 300 |
| QY | 101 | GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro | 120 |
| Db | 301 | GAATTCACCGAGAGCAAGCCCGCCCTGGAGCTACTACCGGATACCTTGCTGCGCCCA | 360 |
| QY | 121 | SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer | 140 |
| Db | 361 | TCACCCCTCTGCGCCACTGGGACGCGAGTGGTCTGAAGACCTTACCAGGACCGTAGC | 420 |
| QY | 141 | LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu | 160 |
| Db | 421 | CTGGCAGGCAAGCTGGAGCCTGTGTCTCCCTCCAGTCCCGCACGCTGACCTGAGCTA | 480 |
| QY | 161 | GluLeuValProProArgLeuSerLysGluLeuIleGlnAsn---MetAspArgVal | 179 |
| Db | 481 | GAGCTGGCGCCATCTCGACTGTCCAAGGAGGAGCTGATCCAGAACAGATTTGGACCCG | 540 |
| QY | 180 | AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGln | 199 |
| Db | 541 | GACCGTGAGATCACCATGTGTAGACGACAGATCTCCAGCTGAAGAGAGAGAGACAC | 600 |
| QY | 200 | LeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProProIle | 219 |
| Db | 601 | TTGGAGGAGGAGCGCGCCCAAGCCCGCCGAAACCCGAGAGCCTGTGTGCGCCACCCATA | 660 |
| QY | 220 | GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu | 239 |
| Db | 661 | GAATCAACACCGAGAGCCTGTGTCCAGATCATCTACGATGAGAACCCGGAAGAGCGGAA | 720 |
| QY | 240 | AlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGln | 259 |
| Db | 721 | GCGGCACCGGATCTTAGAGGCTTGGGGCCCCAGGTGGAGTGGCTCTGTGTACACAC | 780 |
| QY | 260 | ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLys | 279 |
| Db | 781 | CGCTCTGCACACCGCCAGTACCATGAAACATCAAAATAAACCCAGCGCATGCGGAAG | 840 |
| QY | 280 | LeuIleLeuTyrPhenylsArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys | 299 |
| Db | 841 | CTGATCTTTACTTTTAAGCGGAGGAACACCGCGCGCAAGCAGTGGGAACACCTCTTC | 900 |
| QY | 300 | GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn | 319 |
| Db | 901 | CAGCGCTATGACGAGCTCATGGAGGCGCTGGGAGAGAGGTAGAGCGCATAGAACAT | 960 |

Db 5322 CGACCTGTCCAAAGTGCCACACCTGCTGCTGCTGCCGCCGCCGACACAGGCGACCCCGAGC 5381
Qy 1752 aThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHi 1772
Db 5382 CACCGCATGGACCGCTTGTCTTACCTTCCCAACCGCGCCCGAGCCCTTTCAGAGCGGCCCA 5441
Qy 1772 sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrSe 1792
Db 5442 CAGCAGCTCCCACTCTCTCCAGGAGGTCCAAACACTTGCACAAACCAACACACAGTC 5501
Qy 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgGlu 1812
Db 5502 CTGCTCCGAGCGGAGCGAGACCGGATCGAGAGCGGACCGGATCGGAGCGGAGAA 5561
Qy 1812 sSerIleLeuThrSerThrThrValGluHisAlaProIleThrArgProGlyThrG 1832
Db 5562 GTCATCTCTCAGTCCACACGAGCGGTGGAGACGACACCATCTGGAGACTCTGGTACAGA 5621
Qy 1832 uGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlyGlySerSerSerArgProAl 1852
Db 5622 GCAGAGCAGCGGAGCAGCGGAGCAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5681
Qy 1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuG 1872
Db 5682 CTCCCACTCCCGACCCACCGACACTCGCCCATCTCCCTCGGACCCAGGATGCCCTCCA 5741
Qy 1872 nGlnArgProSerValLeuHisThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
Db 5742 GCAGAGACCCAGTGTCTTCCAAACACAGGACATGAAGGGTATCATCACCGCTGTGGAGCC 5801
Qy 1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaTh 1912
Db 5802 CAGCAGCCACCGTCTCGAGGTCCACCTCCCTCACCGTTCGCCAGCTGCCAC 5861
Qy 1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
Db 5862 ATTCCCACTGCCACCACTGCCACTTGGCGGCGACCTCGATGGGTCTTACCTTACCCT 5921
Qy 1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
Db 5922 CATGGAGCCCGTCTTGCTGCCAAGAGAGGCCCCCGGGTCCGCCGCGCAGAGCGGCCCG 5981
Qy 1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
Db 5982 AGCAGACACCGGCATGCCTTCTTCCCAAGCCCCCAGCCGCTCGGGCTGGAGCCCGC 6041
Qy 1972 aSerSerProSerLysGlySerGluProArgProLeuValProValSerGlyHisAl 1992
Db 6042 CTCTCTCCCGCAGCAAGGGCTCGGAGCCCCCGGCCCTAGTGCCTCTCTCTGGCCACGC 6101
Qy 1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr 2012
Db 6102 CACCATCGCCCGCACCCCTCGAAGAACCTTCGACCTCCACCGCAGCCCGGACCCGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe 2032
Db 6162 GGGCCACCTGCTCGGCTCGACCGCCGACCGGGAAAGACTCAAGATAAACCTTTTC 6221
Qy 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
Db 6222 CATCCAGGAACCTGGAACTCCGTTCTCTGGGTACCACCGGCGAGCAGCTACAGCCCGAAGG 6281
Qy 2052 yValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeuProLy 2072
Db 6282 GGTGGAGCCCGCTCAGCCCTGTGAGCTACCCAGTCTGACCCAGCAAGGGGCTCCCCAA 6341
Qy 2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGl 2092
Db 6342 GCACCTGGAAGAGCTCGACAGAGCCACCTGGAGGGGAGCTCGCGCCCAAGCAGCAGG 6401
Qy 2092 yProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSe 2112
Db 6402 CCCCGTGAAGCTTGGCGGGGAGCGCCCACTCCACACCTCGCGCGCGTGTGCTGAGAG 6461

Qy 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
Db 6462 CCAGCCCTCTGTCAGCCCGCTGCTCCAGACCGCCCGAGGGGTCAAAGGTCAACAGCGGGT 6521
Qy 2132 lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPr 2152
Db 6522 GGTCAACCTGGCCCGACACATCAGTGAAGTTCATCACAGAGACTACACCGCGCACCAACC 6581
Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 ACAGCAGCTCAGCGCACCCCTGCCCGCCCTCTACTCTCTCTGCGGCGAGCTGCC 6641
Qy 2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAl 2192
Db 6642 CGTCTGGACCTCGCGCGCCACCCAGTGACCTTACCTCCCGCCCGGACCATGTGTGC 6701
Qy 2192 aProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysTh 2212
Db 6702 CCCGCGCGGTGGTCTCCCGCCACAGCGAAGGGGCAAGAGGTCTCCAGAGCCAAACAAGAC 6761
Qy 2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProGluGlyMetTh 2232
Db 6762 GTCGCTTGGTGGTGGTGGAGACGGTATTGAACCTGTGTCCCCACCGGAGGCGATGAC 6821
Qy 2232 rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh 2252
Db 6822 GGAGCCAGGGCACTCCCGGAGTGTGTGTACCCGCTGTGTACCGGATGGGGAACAGAC 6881
Qy 2252 rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh 2272
Db 6882 GGAGCCCGCAGGATGGGTCTCCAGTCTCCAGGCAACACCCAGCCAGCCGCGCAGCTTCTT 6941
Qy 2272 eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluLeuLeuLy 2292
Db 6942 CAGCAAGTGCACCGGAGCAACTCCGCCATGGTCAAGTCCCAAGAGCAAGAGATCAACA 7001
Qy 2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGl 2312
Db 7002 GAAGCTGAACACCCCAACCGGAATGAGCTGAAATACAATATCAGCCAGCTGGGCGGA 7061
Qy 2312 uIlePheAsnMetProAlaIleThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATGCCCGCCATCACCGGACAGGCTTATGACCTATAGAAGCCAGCGGT 7121
Qy 2332 lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGl 2352
Db 7122 GCAGGAACATGCCAGCAACCAATGGGGCTGGAGGCGCATAAATTAGAAAGGCACTCATGG 7181
Qy 2352 yLysTyrAspGlnTrpGluLeuSerProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACAGTGGGAAGAGTCCCGCGCTCAGCGCCAAATGCTTTTAAACCTCTGAA 7241
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHi 2392
Db 7242 TGCCAGTGCAGCGCTCGCGCTGCTATGCCATAACCGCTGTGTGACGGAGGAGTACCA 7301
Qy 2392 sThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerAr 2412
Db 7302 CACACTCCTCGCCAGGTGGCGGCGGAAGGCAAGGTCTCTGGCAGACCCAGCAGCGG 7361
Qy 2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
Db 7362 AAAGCAAAGTCCCGCGCCCGCGGCTTGGATCTGGGGACCGGCCACCTCTGTCTCTC 7421
Qy 2432 rValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAs 2452
Db 7422 AGTGCACTCGGAGGAGACTGCAACCGCGGACCGCGCTCACCAACCGCGTGTGGGAGGA 7481
Qy 2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGl 2472
Db 7482 CAGGCCCTCTGCGCAGGTTCCACGCCATTCCCTCAACCCCTCATCATGCGGCTGCA 7541

QY 1020 gGlyIysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Db 3162 GGGCAAGAGCAGGAGCGCGCCACCCCGCCGACAGAGGAGGAGAACGCTGTGTCTT 3221
QY 1034 ----AlaPheAlaAlaGluAlaGlnIysLeuProGlyAspProProCysTrpThrSerGI 1052
Db 3222 CCCAGCCTTCGAGCGCGAGCGCCAGAAAGCTGCTGGGGACCCCTTGTGTGACTTCCGG 3281
QY 1052 yLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPr 1072
Db 3282 CTTGCCCTTCCCGTGCCCGCCCGTGAGGTATCAAGGCTCCCGCATGTCGCCGACCC 3341
QY 1072 oSerAlaPheSerTyAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAl 1092
Db 3342 CTCAGCCTTCTCTAGCTCCACTGGTCAACCCACTGCGCCCTGGGCTCCATGACACTGC 3401
QY 1092 aArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl 1112
Db 3402 CCGGCCGCTCTGCGCGGCCACCCACCATCTCAACCCGCTTCCCTCATCTCTCTGCG 3461
QY 1112 aLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGI 1132
Db 3462 CAAGCACCCAGCGTCTTCGAGAGGCAATAGTGCATCTCCCAAGGAATGTGGTCCA 3521
QY 1132 nLeuHisValProTySerGluHisAlaLysAlaProValGlyProValThrMetGlyLe 1152
Db 3522 GCTCCACGTCCTGCTCAGAGCATGCCAAGGCCCGGTGGGCTGTCAACATGGGGCT 3581
QY 1152 uProLeuProMetAspProLysIysLeuAlaProPheSerGlyValIysGlnGluGlnLe 1172
Db 3582 GGCCTTGCCTAGTACCCCAAAAGCTGGCACCTTCAGCGGAGTGAAGAGGAGCGCT 3641
QY 1172 uSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAl 1192
Db 3642 GTCCCAACGGGCGCAGCTGGGCCACCGAGAGCTGGGGTGCCACAGGCCAGAGGC 3701
QY 1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrIysGlyI 1212
Db 3702 GTCCGTGCTGAGAGGACAGCTCTGGCTCAGTTCGGGCGGGAAGCATCACAAAGGCAT 3761
QY 1212 eProSerThrArgValProSerAspSerAlaIleThrTyArgGlySerIleThrHisGI 1232
Db 3762 TCCACGACACGGGTGCCCTCGGACAGCGCCATCATACCGGGCTTCCATACCCACGG 3821
QY 1232 yThrProAlaAspValLeuTyIysGlyThrIleThrArgIleIleGlyGluAspSerPr 1252
Db 3822 CAGCCAGCTGAGTCTCTACAAAGGCGCACCATCACAGGATCATCGGCGGAGACAGCC 3881
QY 1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyroGluI 1272
Db 3882 GAGTCGCTTGGACCGCGCGGAGGACAGCTTGCCTCAAGGGCCAGCTCATCTACGAAGG 3941
QY 1272 yLysIysGlyHisValLeuSerTyroGluGlyMetSerValThrGlnCysSerIysGI 1292
Db 3942 CAAGAAGGGCCAGCTCTGTCTATGAGGTGCGATGCTGTGACCCAGTGTCTCAAGGA 4001
QY 1292 uAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyAs 1312
Db 4002 GGAAGCGCAGAGCAGCTCAGGACCCCGCCATGAGAGCGCGCCCGCCCAAGGCACTATGA 4061
QY 1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGI 1332
Db 4062 CATGATGAGGGCGCGGTGGGACAGCATCTCTCAGCGCAGCATCGAAGGTCTCATGGG 4121
QY 1332 yArgAlaIleProProGluArgHisSerProHisHisLeuIysGluGlnHisIleArg 1352
Db 4122 CCGTGCCATCCCGCGGAGCGACACAGCCCCCACCACCTCAAGAGAGGAGCACCACATCCG 4181
QY 1352 gGlySerIleThrGlnGlyIleProArgSerTyroValGluAlaGlnGluAspTyroLeuAr 1372
Db 4182 CCGGTTCATCACACAGGGATCTCTGCTCTACGTGGAGGACACAGGAGGACTACCTGCG 4241
QY 1372 gArgGluAlaLysLeuLysArgGluGlyThrProProProProProSerArgAs 1392

4242 TCGGAGGCGCAAGCTCCTAAAGCGGAGGCGACGCTCCGCGCCCAACCCCTCAGCGGA 4301
QY 1392 pLeuThrGluAlaTyroLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGI 1412
Db 4302 CTTGACCGAGGCTTACAGACGCGCCCTGGGCCCCCTGAAGCTGAAGCGCGCCATGA 4361
QY 1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAspArgGluI 1432
Db 4362 GGGCTTGTGGTGGCACCGTGAAGAGGCGGCGCTCATCATAGATCCCGCGCAGGA 4421
QY 1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh 1452
Db 4422 GCTGGGCGCACGCGCGAGCTGCCCTGGCCCGCGCGCGCTCAAGAGGCGCTCCATCAC 4481
QY 1452 rGlnGlyThrProLeuLysTyroAspThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
Db 4482 GCAGGCGCACCCGCTCAAGTACGACACCGCGCGTCCACCACTGGCTCCAAAAGCACGA 4541
QY 1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa 1492
Db 4542 CGTACGCTCCCTCATCGGAGCGCGCGCGACGTTCCACCGCTGCACCCCTGTGATGT 4601
QY 1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyroGluGluSerLeuLysSerArgPr 1512
Db 4602 GATGCGCGACGCGCGGCTGGAACGTGCTGTACGAGGAGAGCTTGAAGAGCGCGCC 4661
QY 1512 oGlyThrAlaSerSerSerGlyIysIleAlaArgGlyAlaProValIleValProGI 1532
Db 4662 AGGACCGCGCAGAGCTCGGGGGCTTCATTGGCGCGCGCGCTCATTTGTGCTGA 4721
QY 1532 uLeuGlyIysProArgGlnSerProLeuThrTyroGluAspHisGlyAlaProPheAlaGI 1552
Db 4722 GCTGGGTAAGCGCGCAGAGCGCCCTGACCTATGAGGACACACCGGCGACCTTTCCGG 4781
QY 1552 yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluI 1572
Db 4782 CCACCTCCACAGGTTCCCGGTGACCATGGGAGGCGCACCGCGCTTGAGAGGG 4841
QY 1572 ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluI 1592
Db 4842 CAGCCTTTCGTCAGCAGAGCATCCAGGACCGAAGCTGACGTCCAGCCTCGTGAGAT 4901
QY 1592 eAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyroGI 1612
Db 4902 CGCAAGTCCCGCACAGCAGCGTGGCCGAGCACCAACCCACATCTCGCCTATGA 4961
QY 1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyroArgSerHisIleProLeuAlaph 1632
Db 4962 GCACCTGCTTGGGGCGTGAGTGGCGTGACCTGTATCGAGGCCATCCCTCTGGCCTT 5021
QY 1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyroTyroLeuPr 1652
Db 5022 CGACCCCACTCCATACCCCGCGCATCCCTCTGGAGCGAGCGCTGCTACTACCTGCC 5081
QY 1652 oArgHisLeuAlaProAsnProThrTyroProHisLeuTyroProProTyroLeuIleArgGI 1672
Db 5082 CCGACACCTGGGCCCCCAACCCACCTACCGCACCTGTATCCCACTCCCTCATCCGCGG 5141
QY 1672 yTyroProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyroIleThrSe 1692
Db 5142 CTACCCCGACAGCGCGCGCTGGAGAACCGGACAGCATCATCAATGACTACATCACTC 5201
QY 1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGI 1712
Db 5202 GCAGCAGATGCACACACAGCGCCACCGCCATGGCCGAGGCTGATATGTGAGGGG 5261
QY 1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyroAlaAlaGlyProAspArgGlyIleI 1732
Db 5262 CTTCTCGCCCGCGAGTCTCGCTGCTGCACTCAACTACGCTGCGGTCCCGAGGACATCAT 5321
QY 1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGIyThrProAl 1752

QY 2432 rValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAs 2452
DB 7266 AGTGCACTCGAGGAGAGACTGCAACCGCGGACCGCGCTCACCAACCGCGTGTGGGAGGA 7325
QY 2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLeuMetArgLeuG1 2472
DB 7326 CAGGCCCTCGTCGGCAGGTTCACGCGCATTCGCCCTACAACCCCTCATCGCGGTGCA 7385
QY 2472 nAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
DB 7386 GCGGGGTGTCATGGCTTCCACCCACCGGCGCTCCCGCGGGGAGGGGGCCCTCGC 7445
QY 2492 aGlyProHisHisAlaTrpAspGluProPlysProLeuLeuCysSerGlnTrpGluTh 2512
DB 7446 TGGCCCCCACCACCGCTGGGACGAGGAGGCCCAAGCCACTGCTCTGCTCGCAGTACGAGAC 7505
QY 2512 rLeuSerAspSerGlu 2517.
DB 7506 ACTCTCCGACGCGAG 7521
RESULT 3
ACA62249
ID ACA62249 standard; cDNA; 8686 BP.
XX
AC ACA62249;
DT 12-AUG-2003 (first entry)
XX
DE cDNA encoding human nuclear receptor corepressor SMRte.
KW Human; ss; gene; SMRte; nuclear receptor corepressor; gene therapy;
KW tissue typing; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 157..7680
FT FT /*tag= a
FT FT /product= "SMRte"
XX
PN US2003027137-A1.
XX
PD 06-FEB-2003.
XX
PF 27-MAR-2001; 2001US-00819104.
XX
PR 29-MAR-2000; 2000US-0193138P.
XX
PA (CHEN/) CHEN J D.
XX
PI Chen JD;
XX
DR WPI; 2003-466139/44.
DR P-PSDB; ABU61812.
XX
PT New SMRte proteins and nucleic acids, useful in gene therapy, predictive
PT medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT tissue typing and in forensic biology.
XX
PS Claim 2; Page 32-41; 90pp; English.
XX
CC The invention relates to an isolated SMRte nucleic acid molecule. The
CC nucleic acids are useful in gene therapy, as hybridisation probes for
CC identifying SMRte-encoding nucleic acid molecules and as primers for
CC amplifying of SMRte nucleic acid molecules. The polypeptides are useful
CC as immunogens to raise anti-SMRte antibodies. The SMRte molecules are
CC useful as targets for discovering and developing modulating agents to
CC regulate a variety of cellular processes, in screening assays, in
CC predictive medicine, in therapeutic or prophylactic treatment, in
CC chromosome mapping, tissue typing and in forensic identification of a
CC biological sample. Modulators of SMRte are useful for treating or
CC preventing a condition associated with aberrant SMRte protein or nucleic

CC acid expression or activity, such as cancer. The present sequence
CC represents cDNA encoding the human nuclear receptor corepressor SMRte
XX
SQ Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 7 Gaps: 3

US-09-522-753-5 (1-2517) x ACA62249 (1-8686)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTrpPro 20
DB 157 ATGTGGGGTCCACACAGCGCTGTGGCAGAGAGTGGAGGGCCACTGAGCCCGCTACCG 216
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 217 CCCACAGCGCTTCTTACCCAGTGCAGATCGCCCGGACGACACGAGCGTGGGCTCTG 276
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 277 GAGTACCAGCACCACTCCCGGACTATGCTCTCCACTGTGCGCCCGGCTCCATCATCCAG 336
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 337 CCCCAGCGCGGAGGCCCTCCCTGCTGTGTAGTCTCCAGCCCGGGAATGACGGTCCAG 396
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
DB 397 GAGCTCCACTCGCGCCAGAGTCCCACTCATACCTGCCGAGCTGGGGAAGTCAGAGATG 456
QY 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
DB 457 GAGTTTCATTGAAAGCAAGCGCCCTCGCTAGAGCTGTCTGCTGACCCCTGCTGCCACCG 516
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 517 TCACCCCTGCTGGCCACGGGCCAGCTCGGGATCTGAAGACCTCACCAGGACCGCTAGC 576
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeu 160
DB 577 CTGAGGGGCAAGCTGGAACCGGTGTCTCCCCCAGCCCGCCGACACTGACCTGAGCTG 636
QY 161 GluLeuValProProArgLeuSerLysGluLeuLeuIleGlnAsnMetAspArgValAsp 180
DB 637 GAGCTGGTCCCGCCACGGCTGTCCAAGGAGGAGCTGATCCAGAACATGGACCGCGTGAC 696
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLysLysLysGlnGlnGlnLeu 200
DB 697 CGAGAGATCACCATGTGTAGAGCAGCAGATCTCTAAGCTGAGAGAGAGCAGCAGCTG 756
QY 201 GluGluGluAlaAlaLysProGluProGluLysProValSerProProPheIleGlu 220
DB 757 GAGGAGGAGGCTGCCAAGCGCCGCGGAGCTGAGAAGCCCGTGTCAACCGCCCATCGAG 816
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
DB 817 TCGAAGCAGCGCAGCTGTGTGCAGATCATCTACGACGAGAACCGGAGAGGCTGAAGCT 876
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 877 GCATATCGGATTTCTGAAGGCCCTGGGGGCCCGCCAGGTGAGCTGCGCTGTATCAACACGCCC 936
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAlaMetArgLysLysLeu 280
DB 937 TCCGACACCGGAGTATCATGAGACATCAAAATNAACAGCGGATCGGGAAGAGCTA 996
QY 281 IleLeuTyrPheLysArgArgHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300

| | | | |
|----|------|---|------|
| Db | 5046 | GCACGACATGCACACAAACGCGCCATCGCCAGCGATCATATGCTAGGGG | 5105 |
| Qy | 1712 | YLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAalaglyProArgGlyIleIl | 1732 |
| Db | 5106 | CCTCTCGCCCGCGAGTCTCGCTGGCACTCACTACGCTCGGGTCCCCGAGGCATCAT | 5165 |
| Qy | 1732 | eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl | 1752 |
| Db | 5166 | CGACCTGTCCCAAGTGCACACCTGCCTGTGCTGTCGCCCGACACCGACCCCGCAGC | 5225 |
| Qy | 1752 | aThrAlaMetAspArgIleuAlaTyrLeuProThrAlaProGlnProProSerSerArgHi | 1772 |
| Db | 5226 | CACCGCATGACCGCCTTGCTTACCTCCCGACCGCGCCAGCCCTTCAGCAGCGCCA | 5285 |
| Qy | 1772 | sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrSe | 1792 |
| Db | 5286 | CAGCAGTCTCCCATCTCTCCCGAGAGTCCAAACACTTGACAAAACCAACCAACGCTC | 5345 |
| Qy | 1792 | rSerSerGluArgGluArgAspArgGluArgAspArgGluArgGluArgGluLy | 1812 |
| Db | 5346 | CTCGTCCGAGCGGAGCGAGACCGGATCGAGCGGGAACGGGATCGGAGCGGGAAAA | 5405 |
| Qy | 1812 | sSerIleLeuThrSerThrThrValGluHisAlaProIleIrpArgProGlyThrGl | 1832 |
| Db | 5406 | GTCCATCTCACGTCCACCGACGCGTGGAGCAGCACCCATCTCTGGAGACCTGGTACAGA | 5465 |
| Qy | 1832 | uGlnSerSerGlySerSerGlySerGlyGlyGlyGlyGlySerSerSerArgProAl | 1852 |
| Db | 5466 | GCAGAGCAGCGGACGACGCGCAGCAGCGGGGGTGGGGGACGACGACGCGCCCGC | 5525 |
| Qy | 1852 | aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGl | 1872 |
| Db | 5526 | CTCCCATCTCCATGCCACCCAGCACTCGCCCATCTCCCTTCGGACCCAGAGTGCCTCCA | 5585 |
| Qy | 1872 | nGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr | 1892 |
| Db | 5586 | GCAGAGACCCAGTGTGTTCAACACAGGCATGAAGGGTATCATCACCGCTGTGGAGCC | 5645 |
| Qy | 1892 | oSerLysProThrValLeuArgSerThrSerSerProValArgProAlaAlaTh | 1912 |
| Db | 5646 | CAGCAACCCACCGCTCTGAGGTCCACTCCACTCTCAACCGGTTCGCCACGTGCCAC | 5705 |
| Qy | 1912 | rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe | 1932 |
| Db | 5706 | ATTCTCCACTGCCACCCACTGCCCACTGGGGGACACCTCGATGGGTCTACCTACCT | 5765 |
| Qy | 1932 | uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr | 1952 |
| Db | 5766 | CATGGACCGCTCTGTGTCGAAGAGGCCCCCGGGTCCCGCGCCAGAGCGGCCCG | 5825 |
| Qy | 1952 | gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl | 1972 |
| Db | 5826 | AGCAGACACCGGCATGCTCTCTCGCAAGCCCCACGCCCGCTCCGGGTCTGAGCCCGC | 5885 |
| Qy | 1972 | aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl | 1992 |
| Db | 5886 | CTCTCTCCCGACGAGGGCTCGAGCCCCCGGCCCTAGTGCTCTCTGTCTGTGGCCAGCC | 5945 |
| Qy | 1992 | aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAlaSerProAspProPr | 2012 |
| Db | 5946 | CACCATCGCCGACCCCTCGAAGAACCTCGCACCTCACCGCCAGCCCGGACCCGCC | 6005 |
| Qy | 2012 | oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe | 2032 |
| Db | 6006 | GGCGCCACTGCTCTCGGCTCTGGACCCGCAACCGGAAAAAGACTCAAAAGTAAACCTTTTC | 6065 |
| Qy | 2032 | rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGl | 2052 |
| Db | 6066 | CATCCAGAACTGGAACTCGTTCTCTGGGTATTACCGCAGCAGCACTACAGCCCCGAGG | 6125 |
| Qy | 2052 | yValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeuProLy | 2072 |

| | | |
|------|--|------|
| 6126 | GGTGGAGCCCGTGCAGCCCTGTGAGCTCACCCAGTCTTGACCCACGACGAAGGGGCTTCCCCAA | 6188 |
| 2072 | sHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG | 2092 |
| 6186 | GCACCTGGAAGACTCGACAAGAGCCACCTGGAGGGGGAGCTGGGGCCCAAGCAGCCAGG | 6245 |
| 2092 | yProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSe | 2112 |
| 6246 | CCCCGTGAAGCTTGGCGGGAGGCGGCCACCTCCACACACTCGCGGCGCTGCTCGTAGAG | 6305 |
| 2112 | rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa | 2132 |
| 6306 | CCAGCCCTCGTCAGCCCGCTGTCTCCAGACCGGCCCAAGGGGTCAAAAGGTCAACAGCGGGT | 6365 |
| 2132 | lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPr | 2152 |
| 6366 | GGTCACCTTGGCCAGACACATCATGTAGGTGTCATCACAGAGCTATACCCGGGCACACACC | 6425 |
| 2152 | oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr | 2172 |
| 6426 | ACAGCAGCTCAGCGCACCCCTGCCGCCCCCTACTCTCTCTCCCTGGGCGCAGCTGCC | 6485 |
| 2172 | oValLeuAspLeuArgProProSerAspLeuTyrLeuProProProAspHisGlyAl | 2192 |
| 6486 | CGTCCTGGAGCTCCGCCGCCACCCAGTACCTCTACTCCGCCGCCCGGACCATGTGTGC | 6545 |
| 2192 | aProAlaArgLysProHisSerGluGlyGlyLysArgSerProGluProAsnLysTh | 2212 |
| 6546 | CCGGCCCCGTGGCTCCCCCACAGCAGAGGGGCAAGAGTCTCCAGAGCCAAACAAGAC | 6605 |
| 2212 | rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetTh | 2232 |
| 6606 | GTCCGTCTTGGTGGTGTCAGACCGTATTGAACCTGTGTCCCCACCGGAGGCGCATGAC | 6665 |
| 2232 | rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh | 2252 |
| 6666 | GGAGCCAGGGCACTCCCGGAGTGTGTGTACCCGGCTGTCTACCGGGATGGGAAACAGAC | 6725 |
| 2252 | rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh | 2272 |
| 6726 | GGAGCCAGCAGAGATGGGCTCCAAGTCTCCAGGCACACCAGCCAGCCGCGCATGCTTCTT | 6785 |
| 2272 | eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluLeAsnLy | 2292 |
| 6786 | CAGCAAGCTCAGCCGAGAGCAACTCCGCCATGGTCAAGTCCCAAGAAGCAAGAGATCAACAA | 6845 |
| 2292 | sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGl | 2312 |
| 6846 | GAAAGCTGAACACCCACCAACATGGGCTGAGCCCTGAATACAATATCAGCCAGCCTGGGACGGA | 6905 |
| 2312 | uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa | 2332 |
| 6906 | GATCTTCAAATATGCCGCCCATCACCGGAACAGGCTTATGACCTATATAGAAGCCAGGCGGT | 6965 |
| 2332 | lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGl | 2352 |
| 6966 | GCAGGAACATGCCAGCACCAACATGGGCTGGAGGCCATAATTAGAAAGGCACCTCATGGG | 7025 |
| 2352 | yLysTyrAspGlnTrpGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs | 2372 |
| 7026 | TAAATATGACAGTGGGAGAGTCCCGCGCGCTCAGCGCCAAATGCTTTTAAACCTCTGAA | 7085 |
| 2372 | nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHi | 2392 |
| 7086 | TGCCAGTGCAGGCTGCCCGCTGCTATGCCCATTAACCGCTGCTGACGGAGCGGAGTGACCA | 7145 |
| 2392 | sThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerAr | 2412 |
| 7146 | CACACTCACCTCGCCAGTGGCGGGAGGCCAAGGTCTCTTGGCAGACCCAGCAGCCG | 7205 |
| 2412 | gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe | 2432 |
| 7206 | AAAAGCCAAGTCCCGGCCCGGCTGGATCTTGGGGACCGGCACACCTCTGTCTCTCTC | 7265 |

QY 980 sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProAlaProPr 1000
DB 2886 AGTCCATAGGCCCCCGGAGGACGAGCTCCACCAAGCCAGCTCCCGGAGCCACC 2945
QY 1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
DB 2946 GCCACGGCAAAACCTCAGCGCGAGAGCGACGCCCTCAGCAGCTGGCAGCAGCCCG 3005
QY 1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
DB 3006 GGGCAAGAGCAGGAGCGCGGACCCCGCGCGCAAGAGGAGGAGAGAGCGCTGTGTCTT 3065
QY 1034 ----AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGl 1052
DB 3066 CCCAGCTTCGACAGCGAGGCCAGAGCTGCTGGGAGCCCGCTTGTGGACTTCGG 3125
QY 1052 yLeuProPheProValProProArgGluValLysAlaSerProHisAlaProAspPr 1072
DB 3126 CTGCGCTTCGCGGCGCGCGGCGTGAAGGTGATCAAGGCGCTCCCGCATGCGCGGAGCC 3185
QY 1072 oSerAlaPheSerTyAlaProProGlyHisProLeuProLeuGlyLysHisAspThrAl 1092
DB 3186 CTGAGCTTCTCTAGCTCCACTGGTCAACCGCTGCGCGCTGGGCTCCATGACACTGC 3245
QY 1092 aArgProValLeuProProArgProProThrLysSerAsnProProProLeuLysSerSerAl 1112
DB 3246 CCGGCGCGCTGCGCGGCGCCACCACTCTCAACCGCGCTCCCTCTCTCTCTGCG 3305
QY 1112 aLysHisProSerValLeuGluArgGlnLysGlyAlaLysSerGlnGlyMetSerValGl 1132
DB 3306 CAAGCACCCCGAGGCTCTCGAGAGGCAATAGGTGCGCATCTCCAAAGGAATGTGCGTCCA 3365
QY 1132 nLeuHisValProTySerGluHisAlaLysAlaProValGlyProValThrMetGlyLe 1152
DB 3366 GCTCCACGTCGCTACTCAGAGCATGCCAAGGCGCGGCTGGGCGCTGTCAACATGGGGCT 3425
QY 1152 uProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLe 1172
DB 3426 GCCCTTGCCTGACGACCCCAAGAGCTGGACCTTCAGCGGAGTGAAGCAGGAGCAGCT 3485
QY 1172 uSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAl 1192
DB 3486 GTCCCGGCGGCGGCGGCGGCGGCGGAGCTGGGCGTGGCGGCGGCGGCGGCGGCGG 3545
QY 1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyLysSerLysLysGlyL 1212
DB 3546 GTCGCTGTGAGAGGACAGCTCTGGGCTCAGTTCGCGGCGGAGCATCACCAAGGCGAT 3605
QY 1212 eProSerThrArgValProSerAspSerAlaLysLeuThrArgGlySerLysLysLys 1232
DB 3606 TCCAGACACGCGGCTCGGACAGCGGCATCATACCGCGGCTCCATACCCACCGG 3665
QY 1232 yThrProAlaAspValLeuTyLysGlyThrLysThrArgLysLysLysLysLysLys 1252
DB 3666 CACGCGAGCTGAGCTCTGTACAGGCGCACCATCACGAGATCATCGGCGGAGCAGCC 3725
QY 1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValLysLysLys 1272
DB 3726 GAGTCGCTTGGACCGCGGCGGAGGAGCAGCTGCGCGGCGGAGGCGGCGGCGGCGG 3785
QY 1272 yLysLysGlyHisValLeuSerTyArgGlyGlyMetSerValThrGlnCysSerLysG 1292
DB 3786 CAAGAAGGGCGAGCTGCTGTGCTATGAGGTGGCGATGCTGTGACCGAGGCTCCAAAGGA 3845
QY 1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyAs 1312
DB 3846 GGACGGCAGAGCAGCTCAGGACCCCGCCCATGAGAGCGCGCGCGCGCGCGCGCGCG 3905
QY 1312 pMetMetGluGlyArgValGlyArgAlaLysSerAlaSerLysLysLysLysLysLys 1332
DB 3906 CATGTGAGGCGCGGCGGCGGAGGAGCATCTCTCAGCGCAGCATCGAAGGTCTCTGCG 3965

QY 1332 yArgAlaLysLeuProProGluArgHisSerProHisLysLeuLysGluGlnHisLysLeAr 1352
DB 3966 CCGTGCATCCCGCGGAGGAGCAGACAGCGCCCGACCACTCAAAGAGCAGCAGCACCATCCG 4025
QY 1352 gGlySerLysLeuThrGlnGlyLysLeuProArgSerTyValGluAlaGlnLysAspTyLysLeAr 1372
DB 4026 CCGGCTTCATCACAAGAGGATCCCTCGGTCTTACGTGGAGGACAGGAGGAGTACCTCGG 4085
QY 1372 sArgGluAlaLysLeuLysLeuLysArgGluGlyThrProProProProProProSerArgAs 1392
DB 4086 TCGGAGGCGCAAGCTCTTAAGCGGAGGAGGAGCGCTCCGCCCGCCACCGCTTACCGGGA 4145
QY 1392 pLeuThrGluAlaTyLysThrGlnAlaLeuGlyProLeuLysLeuLysLeuLysLeuLysG 1412
DB 4146 CTTGACCGAGGCTTACAAGACGCGAGCGCTGGGCGCGCTGAACTGAAAGCGCGCCATGA 4205
QY 1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerLysLysLysLysLysLysLysLys 1432
DB 4206 GGGCTCTGGTGGCCACGCTGAAGGAGGCGGCGCTCCATCATGAGATCCCGCGCGAGGA 4265
QY 1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerLysLys 1452
DB 4266 GCTGCGGACACCGCCGAGCTGCGCTGGCGCGCGCGCTCAAGAGGAGGCTCCATCAC 4325
QY 1452 rGlnGlyThrProLeuLysTyAspThrGlyAlaSerThrThrGlySerLysLysLysLys 1472
DB 4326 GCAGGCGCGCGCTCAAGTAGCAGACCGCGCGCTCCACCTGGCTCCAAAAAGCAGCA 4385
QY 1472 pValArgSerLeuLysLysSerProGlyArgThrPheProProValHisProLeuAspVa 1492
DB 4386 CGTACGCTCTCTCATCGGAGCGCGCGCGGAGCTTCCACCGCTGCACCGCTGGATGT 4445
QY 1492 lMetAlaAspAlaAspAlaLeuGluArgAlaCysTyTrpGluSerLeuLysSerArgPr 1512
DB 4446 GATGGCGGAGCGCGCGGCGCTGGAACGCTGCTCTAGAGGAGGAGCTGGAAGAGCGCGCC 4505
QY 1512 oGlyThrAlaSerSerSerGlyLysSerLysLysLysLysLysLysLysLysLysLysLys 1532
DB 4506 AGGCGCGCGCGAGCAGCTCGGGGCGCTCCATTCGCGCGCGCGCGCGCTGCTGCTGA 4565
QY 1532 uLeuGlyLysProArgGlnSerProLeuThrTyTrpGluAspHisGlyAlaProPheAlaGl 1552
DB 4566 GCTGGGTAAAGCGCGCGGAGCGCGCTGACCTATAGAGGAGCAGCGGCGCGCGCTTTCGCG 4625
QY 1552 yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlu 1572
DB 4626 CCACCTCCCGAGGTTTCGCGGAGCTGACCTGCGGAGCGCGCGCGCGCTGCGAGGAGG 4685
QY 1572 ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluL 1592
DB 4686 CAGCTTTCTGTCAGCAGGATCCCGAGGCGGAGCTGACCTGCGCGCGCTGCTGAGAT 4745
QY 1592 eAlaLysSerProHisSerThrValProGluHisLysProHisProLysSerProTyTrpGl 1612
DB 4746 CGCAAGTCCCGCGCAGCAGCGCGCGCGCGCTCTGAGCGGAGCGCGCTGCTTACCTGCC 4805
QY 1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyArgSerHisLysLysLysLysLysLys 1632
DB 4806 GCACCTGCTTCGCGGCGGTGAGTGGCGTGTATCGCAGGCGCATCCCGCTGCGCTT 4865
QY 1632 eAspProThrSerLysProArgGlyLysProLeuAspAlaAlaAlaTyTrpTyLysPr 1652
DB 4866 CGACCCCACTCCATACCCCGCGGCTCCCTCTGAGCGGAGCGCGCTGCTTACCTGCC 4925
QY 1652 oArgHisLeuAlaProAsnProThrTyTrpProHisLysLeuTyProProTyTrpLeuLysArgGl 1672
DB 4926 CGGACACCTGGCGCGCGCGCGCGCGCGCTTACCGCGCGCTGCTTACCTGCC 4985
QY 1672 yTyProAspThrAlaAlaLeuGluAsnArgGlnThrLysLysLysLysLysLysLysLys 1692
DB 4986 CTACCCCGGAGCGCGCGCGCTGGAGAACCGGAGACCATCATCAATGACTACATCACTC 5045
QY 1692 rGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGl 1712

Db 721 GCACATCGGATTCGTGAAGGCGCTGGGGCCCGCAGGTGGAGCTGCGCGTGTACAACAGGCC 780
Qy 261 SerAspThrArgGlnTyrHisGluAsnIleAsnIleAsnGlnAlaMetArgLysLysLeu 280
Db 781 TCCGACACCCGGGAGTATCATGAGAACATCAAAATAAACAGGCGATGCGGAGAAAGCTA 840
Qy 281 IleLeuTyrPheLysArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
Db 841 ATCTTGTAATCAAGAGAGGAATCACGCTCGGAAACAAATGGAGAGAAATGTTCTGCCAG 900
Qy 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
Db 901 CGCTATGACCACTCTGAGGCGCTGGAGAGAGAGGTTCGCGAGTACTACGAGAACGATTC 960
Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
Db 961 CGCGCGGGCCCAAGGAGACAAAGTTTCGCGAGTACTACGAGAACGATTCCTCGAGATC 1020
Qy 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db 1021 CGCAAGCAGCGCGAGCTGCGAGGCGCATGCGAG---AGGTTGGCCAGCGGGCAGTGGG 1077
Qy 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
Db 1078 CTGTCCATGTGCGCGCCCGCAGCAGCAGAGGTGTACAGATCATCGATGGCCTCTCA 1137
Qy 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db 1138 GAGCAGGAGAACCTGGGAGAGCAGATCGCCAGCTGGCGCGTATCGCCCGCATGCTGTAC 1197
Qy 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
Db 1198 GAGCTGACAGCAGCGCATCAAGTTTCATCAATGAAACGGGCTTATGGCCGACCCCATG 1257
Qy 421 LysValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPhe 440
Db 1258 AAGGTGTACAAAGACCGCCAGGTTCATGAACATGTGGAGTGAGCAGAGAGAGACCTTC 1317
Qy 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db 1318 CGGGAGAAAGTTTCATGACAGCATCCCAAGAACTTTGGCCCTGATCGCATATTCCTGGAGAG 1377
Qy 461 LysThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db 1378 AAGACAGTGGCTGAGTGGCTCTTATTAATCTGACTAAGAGAAATGAGAACTATPAG 1437
Qy 481 SerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db 1438 AGCCTGGTGAGCGGAGCTATCGCGCGCGCAAGAGCCAGCAGCAACAACAGCAGCAG 1497
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
Db 1498 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1557
Qy 521 AspGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsn 540
Db 1558 GATGAG 1617
Qy 541 AspLysGluAspLeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGlu 560
Db 1618 GACAAGAGAACCTCTCTCAAGGAGAGACAGACACCTCAGGGAGGAGCAACAGCAG 1677
Qy 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
Db 1678 AAGAGGCTGTGGCTCTCAAGAGCCGCAAACTGTCCAACAGCCAGGGAGAGCAACAGGC 1737
Qy 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1738 CGCATACCCGCTCAATGGCTAATGAGCCCAACAGCGAGAGAGGCCATCACCCCCAGCAG 1797
Qy 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTyrThrGluGluMet 620

Db 1798 AGCGCCGAGCTGGCCTCCATGAGAGCTGAATGAGAGTTCTCGCTGGACAGAGAAGAATG 1857
Qy 621 GluThrAlaLysLysGlyLeuLeuHisGlyArgAsnTyrSerAlaIleAlaArgMet 640
Db 1858 GAAACAGCCCAAGAAAGCTCTCTGGAAACAGCGCGCAACTGGTGGCCATCGCCCGGATG 1917
Qy 641 ValClySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660
Db 1918 GTGGCTCCAGACTGTGTGCGAGTGAAGACTTCTACTTCACTACAGAGAGAGCAG 1977
Qy 661 AsnLeuAspGluLeuLeuGlnHisLysLysMetGluLysGluArgAsnAlaArg 680
Db 1978 AACCTCGATGAGATCTTCGACAGCAGCAGCTGAAGATGAGAGAGAGAGAACGCGCG 2037
Qy 681 ArgLysLysLysAlaProAlaAlaAsnGluGluAlaPheProProValVal 700
Db 2038 AGGAAGAGAGAAAGCGCGCGCGCGCAGCAGAGAGGCTGCATTCCTCCCGCGTGGTG 2097
Qy 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
Db 2098 GAGGATGAGGAGATGGAGGCGTGGGGCTGACGGGAATGAGGAGGAGATGGTCGAGGAG 2157
Qy 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db 2158 GCTGAA-----GCC 2166
Qy 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
Db 2167 ACTGTCAACAAACAGCTCAGACACCGAGAGCATCCCTCTCTCTACACTGAGGCGCCCAAG 2226
Qy 761 AspThrGlyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProPro 780
Db 2227 GACACAGGCGAGAATGGCCCAAGCCCGCCAGCCCTGGCGCGCAGCGGCCACCCCA 2286
Qy 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
Db 2287 GGGCCACCCACCCACCCAGCGAGGACATCCCGGCC-CCCACTGAGTCCACCCCGGCTC 2345
Qy 800 rGluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProProPr 820
Db 2346 TGAAGCCACCTTAGCCCTTAGCCCTCAGCCCGCCAGCAGACCCCACTTCCCTTTCACCTCTCC 2405
Qy 820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGl 840
Db 2406 TGTGTGTCCTCCAG 2465
Qy 840 yGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGl 860
Db 2466 GGAGGAGCAGAAAGCCCGCGCTGAGGAGCTGGCAGTGGACACAGAGAGAGCGAGGA 2525
Qy 860 uProValLysSerGluCyethrGluGluAlaGluGluGlyProAlaLysGlyLysAspAl 880
Db 2526 GCCCGTCAAGAGCGAGTGCACGGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 2585
Qy 880 aGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyLys 900
Db 2586 GGAGGCGCTGAGGCCACCGCCAGAGGCGCTCAAGCGCAGAGAGAGAGAGAGAGAGAG 2645
Qy 900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaTh 920
Db 2646 CGGCAGGGCCACACAGCCCAAGAGCTCGGGCGCGCCCGCCAGGACAGCACTCCAGTGCAC 2705
Qy 920 rCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPr 940
Db 2706 CTGAGTGCAGACAGGTGATGAGCGCGGAGCGCGCAGAACCGCTGCTGTCTTCTCC 2765
Qy 940 oArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPr 960
Db 2766 AAGGCCAGCCTCTCTCACCCGACTGGCGAGCCCGCGGCCAAATGCTCTCACCCAGAGACC 2825
Qy 960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLy 980
Db 2826 ACTGGACCTGAAGCAGTGAAGCAGCGCGGTGCCATCCCGCCCATCCCGGTCAGGTCA 2885

```
Db 1141 GCTATGCCATAACCGCTGCTGACGACGAGTGACACACACTCACTCGCCAGGTGGC 7200
Qy 2400 GlyGlyValValSerGlyArgProSerSerArgValValSerProAlaPro 2419
Db 7201 GCGGGAAGCCCAAGGTCCTGGCAGACCCAGCAGCCGAAAGCCCAAGTCCCCGGCCCCG 7260
Qy 2420 GlyLeuAlaSerGlyArgProProSerValSerSerValHisSerGluGlyArgCys 2439
Db 7261 GGCCTGGCATCTGGGACCGCCACCCCTCTGCTCTCACTGACCTGGAGGGAGACTGC 7320
Qy 2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
Db 7321 AACCGCGGACGCGCTCACCAACCGGCTGTGGGAGGACAGGCGCTCGTCCGAGGTTC 7380
Qy 2460 ThrProPheProTrpAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
Db 7381 ACGCCATTCCTTACAAACCCCTGATCATCGGCTGAGCGGGGTGTCATGGCTTCCCCA 7440
Qy 2480 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAsp 2499
Db 7441 CCCCCACCGGCTCCCCCGGGGCGAGCGGCGCCCTCGCTGCCCCCACCACGCTGGGAC 7500
Qy 2500 GluGluProLysProLeuLeuCysSerGlnTrpGluThrLeuSerAspSerGlu 2517
Db 7501 GAGGAGCCCAAGCACTGCTCTGCTCGCAGTAGCAGACACTCTCCGACAGCGAG 7554

RESULT 2
AC62250
ID AC62250 standard; cDNA; 7521 BP.
XX
AC AC62250;
DT 12-AUG-2003 (first entry)
XX
DE Human nuclear receptor corepressor SMRTE coding region cDNA.
XX
KW Human; ss; gene; SMRTE; nuclear receptor corepressor; gene therapy;
KW tissue typing; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..7524
FT /*tag= a
FT /partial
FT /product= "SMRTE"
FT /note= "No stop codon given"

US2003027137-A1.
XX
PN 06-FEB-2003.
XX
PF 27-MAR-2001; 2001US-00819104.
XX
PR 29-MAR-2000; 2000US-0193138P.
XX
PA (CHEN/) CHEN J D.
XX
PI Chen JD;
XX
WP1; 2003-466139/44.
XX
DR P-PSDB; ABU61812.
XX
XX
PT New SMRTE proteins and nucleic acids, useful in gene therapy, predictive
PT medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT tissue typing and in forensic biology.
XX
PS Claim 2; Page 48-56; 90pp; English.
XX
CC The invention relates to an isolated SMRTE nucleic acid molecule. The
CC nucleic acids are useful in gene therapy, as hybridisation probes for
CC identifying SMRTE-encoding nucleic acid molecules and as primers for
```

```
CC amplifying of SMRTE nucleic acid molecules. The polypeptides are useful
CC as immunogens to raise anti-SMRTE antibodies. The SMRTE molecules are
CC useful as targets for discovering and developing modulating agents to
CC regulate a variety of cellular processes, in screening assays, in
CC predictive medicine, in therapeutic or prophylactic treatment, in
CC chromosome mapping, tissue typing and in forensic identification of a
CC biological sample. Modulators of SMRTE are useful for treating or
CC preventing a condition associated with aberrant SMRTE protein or nucleic
CC acid expression or activity, such as cancer. The present sequence
CC represents the human nuclear receptor corepressor SMRTE coding region
CC cDNA
XX
SQ Sequence 7521 BP; 1635 A; 2728 C; 2212 G; 946 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 0 Length: 7521
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 7 Gaps: 3
```

US-09-522-753-5 (1-2517) x ACA62250 (1-7521)

```
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTrpPro 20
Db 1 ATGTCGGGCTCCACACAGGCTGTGGCACAGACGTGGAGGGCCACTGAGCCCGCTACCCG 60
Qy 21 ProHisSerLeuSerTrpProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 61 CCCCACAGCCTTTCTTACCAGTGCAGATCGCCCGGACGACACGACGCTGGGGCTCTG 120
Qy 41 GluTrpGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 121 GAGTACGACACCACTCCCGGACTATGCTCCCACTGTGCGCCCGGCTCCATCATCCAG 180
Qy 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 181 CCCCACGGCGGAGGCGCTCCCTGCTGTCTGAGTTCACGCGCGGAATGAACGGTCCCAG 240
Qy 81 GluLeuHisLeuArgProGluSerHisSerTrpLeuProGluLeuGlyLysSerGluMet 100
Db 241 GAGCTCCACTGCGGCGCAGAGTCCCACTCATACCTCCCGAGTGGGGAAGTCAGAGATG 300
Qy 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 301 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTGCCTGACCCCTGCTCGACCG 360
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 361 TCACCCCTGCTGGCCACGCGGCCAGCTGCGGGATCTGAAGACCTCACCAAGGACCGTAGC 420
Qy 141 LeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeu 160
Db 421 CTGACGGGCAAGCTGGAACCGGTGCTCCCGCCAGCCCGCCGACACTGACCTTGAGCTG 480
Qy 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
Db 481 GAGCTGGTGGCCGACAGCGGTGTCCAAAGGAGGAGCTGATCCAGAACATGGACCGCGTGGAC 540
Qy 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnLeu 200
Db 541 CGAGAGATCACCATGCTAGAGCAGCAGATCTCTTAAGCTGAAGAAGAAGCAGCAACAGCTG 600
Qy 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu 220
Db 601 GAGGAGGAGGCTGCCAAGCGCGCCGAGCTGAGAGCCCGTGCACCGCCCGCCATCGAG 660
Qy 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
Db 661 TCGAAGACCGCAGCTGTGTGACAGATCATCTACACGAGNACCGGAGAGAGGCTGAAGCT 720
Qy 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyAsnGlnPro 260
```


| | | | |
|----|------|---|------|
| Db | 2761 | ACCTGCAGTCAGACGAGGTGGATGAGCGCGCGCACAAAGACCGGCTGCTGCC | 2820 |
| Qy | 940 | ProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys | 959 |
| Db | 2821 | CCAAGGCCAGCTCTCTCACCCGACTGGCGAGCCCCCGGCCCAATGCTCACCCCAAG | 2880 |
| Qy | 960 | ProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaProProGlnValThr | 979 |
| Db | 2881 | CCACTGGACCTGAAGCAGCTGAAGCAGCAGCGGCTGCCATCCCCCATCCAGGTACC | 2940 |
| Qy | 980 | LysValHisGluProProArgGluAspAlaProThrLysProAlaProAlaPro | 999 |
| Db | 2941 | AAAGTCCATGAGCCCCCGCGGAGGACGAGCTCCCAACGAGCAGCTCCCCCCAGCCCCA | 3000 |
| Qy | 1000 | ProProGlnAsnLeuGlnProGluSerAspAlaProGlnProGlnProGlnSerPro | 1019 |
| Db | 3001 | CCGCCACCGCAAACTTCGACGCGGAGAGCGCCCTCAGCAGCCTGGCAGCAGCCCC | 3060 |
| Qy | 1020 | ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla | 1039 |
| Db | 3061 | CGGGGCAAGAGCAGGAGCCCGGCACCCCGCCGACAAAGAGGCTTCGAGCGGAGGCC | 3120 |
| Qy | 1040 | GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro | 1059 |
| Db | 3121 | CAGAAGCTGCTGGGAGCCCCCTTGCTGGACTTCGCGCTGCCCTTCCCGCTGCCCCCC | 3180 |
| Qy | 1060 | ArgGluValLleLysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaPro | 1079 |
| Db | 3181 | CGTGAGGTGATCAAGGCTCCCGCATGCCCCGAGCCCTCAGCCTTCTCTCAGCTCCA | 3240 |
| Qy | 1080 | ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro | 1099 |
| Db | 3241 | CTTGTCACCACTGGCCCTTGGGCTCCATGACACTGCCCGCCGCTCTGCCGCGCCA | 3300 |
| Qy | 1100 | ProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGlu | 1119 |
| Db | 3301 | CCACCATCTCCAACCGCTCCCTCTCATCTCTGCGCAAGCACCCACGCGTCTCGAG | 3360 |
| Qy | 1120 | ArgGlnIleGlyAlaLleSerGlnGlyMetSerValGlnLeuHisValProTyTrSerGlu | 1139 |
| Db | 3361 | AGGCAAAATAGTGTCATCTCCCAAGGAATTCGGTCAGCTCCACGTCGCGTACTCAGAG | 3420 |
| Qy | 1140 | HisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLys | 1159 |
| Db | 3421 | CATGCCAAGCCCCGCTGGGCCCTGTCCACTGGGGCTGCCCTGCCCATGGACCCCAA | 3480 |
| Qy | 1160 | LysLeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGly | 1179 |
| Db | 3481 | AAGCTGGCACCTTCAGCGAGTGAAGCAGGAGCAGCTGTCCCCACGGGGCCAGGCTGGG | 3540 |
| Qy | 1180 | ProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAla | 1199 |
| Db | 3541 | CCACCGGAGAGCTGGGGGTGCCACAGCCACGAGGCGTCCGTGCTGAGAGGACAGCT | 3600 |
| Qy | 1200 | LeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSer | 1219 |
| Db | 3601 | CTGGGCTCAGTTCGGGCGGAAGCATCCCAAGGCATTCACAGCACACGGGHTGCCCTCG | 3660 |
| Qy | 1220 | AspSerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspValLeuTy | 1239 |
| Db | 3661 | GACAGGCCCATCATACCGCGCTCCATCACCCAGCGCACGCTGACGCTCCTGTAC | 3720 |
| Qy | 1240 | LysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArg | 1259 |
| Db | 3721 | AAGGGCACCATCACAGGATCATCGGCGAGGACAGCCGAGTGGCTTGGACCGCGCGCG | 3780 |
| Qy | 1260 | GluAspSerLeuProLysGlyHisValIleTyGluGlyLysLysGlyHisValLeuSer | 1279 |
| Db | 3781 | GAGACAGCTGCCCAAGGCCACGTATCTAGAAGGCAAGAGGCCACAGCTTGTGTC | 3840 |
| Qy | 1280 | TyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGly | 1299 |

| | | | |
|----|------|--|------|
| Db | 3841 | TATGAGGTGGCATGTCTGTGACCCAGTGTCTCCAGGAGGACGGCAGAAAGCAGCTCAGGA | 3900 |
| Qy | 1300 | ProProHisGluThrAlaAlaProLysArgThrTyAspMetMetGluGlyArgValGly | 1319 |
| Db | 3901 | CCCCCCCATGAGAGCGGCCCCCAAGCGCACCTATGACATGATGAGGCGCCGCTGGGC | 3960 |
| Qy | 1320 | ArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg | 1339 |
| Db | 3961 | AGAGCCATCTCTCAGCAGCATCGAAGGTCTCATGGCGCGTGCCATCCCGCCGAGCGGA | 4020 |
| Qy | 1340 | HisSerProHisLysLysGlnHisHisIleArgGlySerIleThrGlnGlyIle | 1359 |
| Db | 4021 | CACAGCCCCCACCATCAAGAGCAGCACCATCCGCGGTCCATCACACAGGGAATC | 4080 |
| Qy | 1360 | ProArgSerTyValGluAlaGlnGluAspTyLeuArgArgGluAlaLysLeuLys | 1379 |
| Db | 4081 | CCTCGGTCTCTAGTGAGGACACAGGAGGACTACTCGCTGGGAGGCCAAGCTCCTAAAG | 4140 |
| Qy | 1380 | ArgGluGlyThrProProProProProProProProProProProProProProProPro | 1399 |
| Db | 4141 | CGGGAGGCGACGCTCCGCCCCACCGCTCAGCGGACCTGACCGAGGCTTACAGACG | 4200 |
| Qy | 1400 | GlnAlaLeuGlyProLeuLysLysProAlaHisGluGlyLeuValAlaThrValLys | 1419 |
| Db | 4201 | CAGGCCCTGGGCCCTGAAAGCTGAAGCGGCGCCATGAGGCGCTGGTGGCCACGGTGAAG | 4260 |
| Qy | 1420 | GluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeu | 1439 |
| Db | 4261 | GAGCGGGCGCTTCCATGATGATCCCGCGGAGGAGCTGGGCGACACGCCCGAGCTG | 4320 |
| Qy | 1440 | ProLeuAlaProArgProLeuLysGlySerIleThrGlnGlyThrProLeuLysTy | 1459 |
| Db | 4321 | CCCTGGCCCCCGCGCGCTCAAGGAGGCTCCATCAGCAGGCGACCCCGCTCAAGTAC | 4380 |
| Qy | 1460 | AspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySer | 1479 |
| Db | 4381 | GACACGGCGCGTCCACCACTGGCTCCAAAAGCACGACGTACGCTCCCTCATCGGCAGC | 4440 |
| Qy | 1480 | ProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeu | 1499 |
| Db | 4441 | CCCGCGCGACGTTCCACCCGCTGACCCGCTGGATGTATGGCCACGCCCGGCGACTG | 4500 |
| Qy | 1500 | GluArgAlaCysTyArgGluSerLeuLysSerArgProGlyThrAlaSerSerSerGly | 1519 |
| Db | 4501 | GAACTGCTCTGTACGAGGAGGAGCTGAGAGCGCGCCAGGACCCCGCAGCAGCTCGGG | 4560 |
| Qy | 1520 | GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSer | 1539 |
| Db | 4561 | GGCTCCATTGGCGCGCGCGCTGGTCAATTGTGCTGAGCTGGGTAAAGCCGCGCAGAGC | 4620 |
| Qy | 1540 | ProLeuThrTyGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro | 1559 |
| Db | 4621 | CCCTTAACCTATGAGACACCGCGGACCCCTTGGCGGCGACCTCCACAGGTTGCGCC | 4680 |
| Qy | 1560 | ValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAla | 1579 |
| Db | 4681 | GTGACCACCGGGAGGCCACCGCGCTTCGAGAGGCGAGCTTCGTCGTCAGCAAGGCA | 4740 |
| Qy | 1580 | SerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThr | 1599 |
| Db | 4741 | TCCAGGACCGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG | 4800 |
| Qy | 1600 | ValProGluHisHisProHisProHisProHisProHisProHisProHisProHisPro | 1619 |
| Db | 4801 | GTGCCCGAGCACCCACACCCCATCTCGCCCTATGAGCACCTGTCTGGGGCGGTGAGT | 4860 |
| Qy | 1620 | GlyValAspLeuTyArgSerHisIleProLeuAlaPheAspProThrSerIleProArg | 1639 |
| Db | 4861 | GGCGTGGACCTGTATCGAGGCCACATCCCTCCCTGGCTTCGACCCCACTCCATCCCGCG | 4920 |
| Qy | 1640 | GlyIleProLeuAspAlaAlaAlaIleTyTrLeuProArgHisLeuAlaProAsnPro | 1659 |
| Db | 4921 | GGCATCCCTCTGAGCAGCGCGCTGCTACTACTCTGCCCGACACCTGGCGCCCCCAACCC | 4980 |

QY 201 GluGluGluAlaAlaLysProGluProGluLysProValSerProProProlleGlu 220
DB 601 GAGGAGGAGCTGCCAAGCGCCGAGCGCTGAGAGGCCGTGTACCGCCGCCCATCGAG 660
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAenArgLysLysAlaGluAla 240
DB 661 TCGAAGCAGCGAGCTGGTGTGAGATCATCTAGCAGAGAACCGAAGAGGCTGAAGCT 720
QY 241 AlaHisArgIleLeuGluGluLysGluProGluValGluLeuProLeuTyrAsnGlnPro 260
DB 721 GCACATCGGATTCTGGAAGGCCCTGGGGCCCGAGCTGGCGCTGTATCAACCGAGCCC 780
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 781 TCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACAGCGCATCGGAGAGAGCTA 840
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
DB 841 ATCTTGTACTCAAGAGAGGAATCACGCTCGGAACAATGGGAGCAGAGAAGTTCTGCCAG 900
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
DB 901 CGCTATGACGAGCTCATGGAAGGCCCTGGCAAAAAAAGGTGGAGCGCATCGAGAAACAACCCC 960
QY 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
DB 961 CGCGCGCGGCCAAGAGACAGAGTGGCGGAGTACTACGAAAGCAGTTCCCTGAGATC 1020
QY 341 ArgLysGlnArgGluLeuGlnArgMetGlnSerArgValGluGlnArgGlySerGly 360
DB 1021 CGCAAGCAGCGAGCTGCAGGAGCGCATCGAGGCGAGGTGGCGCAGCGGGGCGAGTGGG 1080
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1081 CTGTCCATGTGCGCGCCCGCAGCAGCAGAGGTGTGAGATCATCGATGGCGCTCTCA 1140
QY 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1141 GAGCAGGAGAACCTGGGAGAGCAGATGCGCAGCTGGCGCTGATCCGCGCCCATGCTGATC 1200
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1201 GACGCTGACCAGCAGCGCATCAAGTTTCATCAATGAACGGGTATGGCGCAGACCCCATG 1260
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440
DB 1261 AAGGTGTACAAAGACCGCCAGGTGATGAATGTGGAGTGAGCAGGAGAGAGACCTTC 1320
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 1321 CGGGAGAGTTTCATGACGATCCCAAGAACTTTGGCCCTGATCGCATCTTCTGGAGAGG 1380
QY 461 LysThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
DB 1381 AAGACAGTGGCTAGTGCCTCTATTACTATTACTGATGAAGAGATGAGAACTATAAG 1440
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSer---GlnGlnGlnGlnGln 499
DB 1441 AGCCTGTGAGAGAGAGCTATCGCGCGCGGCAAGAGCGCAGCAGCAGCAACACAGCAG 1500
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGlu 519
DB 1501 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCCATGCCCGCAGCAGCAGGAGGAG 1560
QY 520 LysAspGluLysGluLysGluAlaGluLysGluGluLysProGluValGlu 539
DB 1561 AAAGATGAGAGGAGAGAGAGAGAGCGGAGAGAGGAGGAGAGAGCGGAGTGGAG 1620
QY 540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAsnAsp 559
DB 1621 AACGACAAAGAGACCTCTCTCAAGGAGAGACAGACACCTCAGGGGAGGAGCAACGAC 1680

QY 560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
DB 1681 GAGAGGAGGCTGTGGCTCCAAAGGCGCGCAAACTGCCAACAGCGCAGGAGAGCGCAA 1740
QY 580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluAlaIleThrProGln 599
DB 1741 GGCCGATCATCCCGCTCAATGGCTAATAGGCCCAACAGCAGGAGGCGCATCACCCCGAG 1800
QY 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGlu 619
DB 1801 CAGAGCGCGAGCTGGCTCCATGGAGCTGAATGAGAGTTCTCGTGGACAGAGAGAGAA 1860
QY 620 MetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArg 639
DB 1861 ATGGAAACAGCCCAAGAAAGGTCTCTTGGAAACACGGCGCAACTGGTCCGCGCATCGCCGG 1920
QY 640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
DB 1921 ATGGTGGGCTCCAAAGACTGTGTCCGAGCTGTAAAGACTTCTACTTCAACTACAGAGAGG 1980
QY 660 GlnAsnLeuAspGluIleLeuGlnHisLysLysMetGluLysGluArgAsnAla 679
DB 1981 CAGAACCTCGATCGATCTTGGCAGCAGCACAGCTGAAGATGGAGAGGAGAGAGCGG 2040
QY 680 ArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProVal 699
DB 2041 CGGAGGAAGAAAGAAAGCGCGCGCGCGCAGCGAGGAGCTGCATTCCGCGCCCGTG 2100
QY 700 ValGluAspGluIleMetGluAlaSerGlyValSerGlyAsnGluGluMetValGlu 719
DB 2101 GTGGAGATGAGAGATGGAGGCGTGGGCGCTGAGCGGAATCAGGAGAGATGGTGGAG 2160
QY 720 GluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyPro 739
DB 2161 GAGCTGGAAGCTTACATGCTCTGGGAATGAGTGGCCAGAGGGGAATCGATGGGCCCA 2220
QY 740 AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAla 759
DB 2221 GCCACTGTCAACAACAGCTCAGACACCGAGAGCATCCCTCTCTCACAGCGAGCGGCC 2280
QY 760 LysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProPro 779
DB 2281 AAGGACACAGGGCAGAAATGGGCCCAAGCCCGCCAGCCACCTGGCGCGCGAGCGGCCACC 2340
QY 780 ProGlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAla 799
DB 2341 CAGGCGCCACCCACCCACACCGAGGAGCATCCCGGGCCCCCATTTGAGCCACCCCGGCC 2400
QY 800 SerGluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProPro 819
DB 2401 TCTGAAGCCACCGAGCGCCCTACGCGCCCAACACAGCACCCCGCATCGCCCTCTGCACCTCCT 2460
QY 820 ProValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGlu 839
DB 2461 CCTGTGTCTCCCAAGGAGAGAGGAGGAGAGCGCAGCAGCGCCCGCCAGTGGAGAGAG 2520
QY 840 GlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu 859
DB 2521 GGGGAGGAGCAGAAAGCCCCCGCGCTGAGAGCTGGCAGTGGACACAGGAGGAGCGCGAG 2580
QY 860 GluProValLysSerGluCysThrGluGluAlaGluGluProAlaLysGlyLysAsp 879
DB 2581 GAGCGCGCTCAAGAGCGAGTGCAGGAGGAAGCGGAGGCGCGCGCGCAAGGGGCAAGGAC 2640
QY 880 AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGly 899
DB 2641 GCGAGGCGCGCTGAGGCCACGCGCGGAGGCGCGCTCAAGGAGAGAGAGAGGAGGCGGG 2700
QY 900 SerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAla 919
DB 2701 AGCGGAGGCGCCACCTGCAAGAGCTCGGGCGCCCCCGCAGGAGCAGCTCCAGTGCT 2760
QY 920 ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer 939

c 82 606 4.6 34094 8 AAC781130
 83 603.5 4.6 28906 4 AAH23705
 84 601 4.5 38506 7 ABS56090
 85 600 4.5 36778 3 AAZ87318
 86 600 4.5 37948 3 AAZ87285
 87 600 4.5 38506 3 AAA56333
 88 600 4.5 38506 3 AAZ56001
 89 600 4.5 38506 7 ADA09418
 90 598 4.5 110000 2 AAQ9683_06
 91 596 4.5 29879 2 AAQ46806_06
 92 595 4.5 8546 7 ABV75361
 93 593.5 4.5 9115 4 AAH23695
 94 586 4.4 11145 7 ACD13385
 95 586 4.4 12227 7 ACC50291
 96 585.5 4.4 499 8 ACH32484
 97 584.5 4.4 110000 4 AAQ9683_18
 98 583.5 4.4 11680 3 AAC76700_06
 99 581.5 4.4 110000 4 AAQ9682_18
 100 578 4.4 13715 6 ABQ61152

ALIGNMENTS

RESULT 1
 AAC74783
 ID AAC74783 standard; cdna; 8564 BP.
 XX
 AC AAC74783;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF338 polynucleotide sequence SEQ ID NO:675.
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparkinsonian; antiparkinsonian; nontropic; neuroprotective;
 KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; sb.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US008621.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimketa RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR P-PSDB; AAB40574.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 5; Page 779-784; 5507pp; English.

XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparkinsonian; antiparkinsonian; nontropic; neuroprotective; osteopathic;
 CC anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antinaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX
 SQ Sequence 8564 BP; 1859 A; 3034 C; 2532 G; 1139 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 8564
 Score: 13178.50 Matches: 2512
 Percent Similarity: 99.80% Conservative: 1
 Best Local Similarity: 99.76% Mismatches: 4
 Query Match: 99.72% Indels: 1
 DB: Gaps: 1

US-09-522-753-5 (1-2517) x AAC74783 (1-8564)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrPro 20
 DB 1 ATGTCGGGTCCACACAGCTTGGGCACAGACGTGGAGGCCACTGAGCCCGCTACCCG 60
 QY 21 ProHisSerLeuSerTyrProValGlnLeuAlaArgThrHisThrAspValGlyLeuLeu 40
 DB 61 CCCACAGGCTTCTCCACGATGTCGCGGACGACGACGACGACGACGACGACGACGACG 120
 QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleGln 60
 DB 121 GAGTACAGCAGCCACTCCCGGACTATGCTCCACGCTGTCGCGGCTCCATCATCCAG 180
 QY 61 ProGlnArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
 DB 181 CCCACGCGGAGGCGCCCTCCCTGCTGTGAGTTCAGCCCGGGAATGACCGTCCAG 240
 QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
 DB 241 GAGCTCCACTGCGGCGCAGATGCCACTCATCTGCGGAGTGGGAGTCCAGATG 300
 QY 101 GluPheIleGluSerLysArgProArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 120
 DB 301 GAGTTCATTGAAAGCAAGCGCCCTCGCTAGAGTGTGCTGCTGACCCCTGCTGCGACCG 360
 QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
 DB 361 TCACCCCTGCTGCGGCGCAGGCGCTCGGGGATCTGAAGACCTCACCAGGACCGGTAGC 420
 QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
 DB 421 CTGACCGGCGAAGCTGGAAACGGGTGTCTCCCGCAGCCCGCCGACACTGACCTGAGCTG 480
 QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
 DB 481 GAGTGTGTCGGCGCAGCGCTGTCCAAGGAGGAGCTGATCCAGAACATGAGCCGCGTGAC 540
 QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeu 200
 DB 541 CGAGAGATCACCATGTTAGTACGACGACGAGTCTCTTAAGCTGAAGAAGAAGCAACAGCTG 600

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 15, 2004, 22:22:55 ; Search time 1876 Seconds

(without alignments)
5699.741 Million cell updates/sec

Title: US-09-522-753-5

Perfect score: 13215
Sequence: 1 MSGSTQLVQWTRATEPRYP.....WDEPKPLCSQVETLSDSE 2517

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Ygapop 6.0 , Ygapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09522753/runat_15042004_143737_17420/app_query.fasta_1.2695
-DB=N_Geneseq_29Jan04 -QWTF=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=100 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=100
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09522753@cgn_1_1431 @runat_15042004_143737_17420 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*

1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002as: *
7: Geneseqn2003as: *
8: Geneseqn2003bs: *
9: Geneseqn2003cs: *
10: Geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | ID | Description |
|------------|---------|-------|-------|--------|----------|--------------------|
| 1 | 13178.5 | 99.7 | 8564 | 3 | AAC74783 | Aac74783 Human ORF |
| 2 | 12978 | 98.2 | 7521 | 7 | ACA62250 | ACA62250 Human nuc |
| 3 | 12978 | 98.2 | 8686 | 7 | ACA62249 | ACA62249 cDNA enco |
| 4 | 10832.5 | 82.0 | 7386 | 7 | ACA62452 | ACA62452 Mouse nuc |
| 5 | 10832.5 | 82.0 | 8544 | 7 | ACA62451 | ACA62451 cDNA enco |
| 6 | 7885 | 59.7 | 5989 | 6 | ABK84305 | ABK84305 Human cDN |
| 7 | 4612 | 34.9 | 2930 | 3 | ADC35130 | ADC35130 Human bre |
| 8 | 4187 | 31.7 | 7780 | 3 | AAA60630 | AAA60630 HNRCK nuc |

| | | | | | | |
|----|--------|------|--------|---|----------|---------------------|
| 9 | 4147.5 | 31.4 | 7940 | 7 | ABZ34833 | Abz34833 Coding se |
| 10 | 4043 | 30.6 | 7900 | 3 | AAA60629 | AAA60629 Human HNR |
| 11 | 1871.5 | 14.2 | 2336 | 3 | AD31306 | Ad31306 Human dia |
| 12 | 1113 | 8.4 | 650 | 7 | ACA57401 | Acas7401 Human adi |
| 13 | 1083 | 8.2 | 10910 | 4 | ABU03131 | Abu03131 Drosophill |
| 14 | 1076 | 8.1 | 872 | 3 | AAA02670 | Aaa02670 Human col |
| 15 | 960 | 7.3 | 555 | 7 | ACA57524 | Acas7524 Human adi |
| 16 | 880 | 6.7 | 718 | 9 | AD876358 | Ad876358 Human BSK |
| 17 | 845 | 6.4 | 710 | 9 | AAI97539 | Aai97539 Human neu |
| 18 | 774.5 | 5.9 | 673 | 7 | ACA57523 | Acas7523 Human adi |
| 19 | 746.5 | 5.6 | 520 | 6 | ABK44934 | Abk44934 cDNA enco |
| 20 | 726.5 | 5.5 | 33529 | 5 | AAS17367 | Aas17367 DNA seque |
| 21 | 723.5 | 5.5 | 27705 | 9 | ADC26979 | Adc26979 Sorangium |
| 22 | 718 | 5.4 | 1922 | 9 | ADT31931 | Adt31931 Retinoid |
| 23 | 708.5 | 5.4 | 67251 | 9 | ADC26995 | Adc26995 Sorangium |
| 24 | 702 | 5.3 | 20222 | 9 | ADC26981 | Adc26981 Sorangium |
| 25 | 700 | 5.3 | 13416 | 7 | AD55815 | Aad55815 Micromono |
| 26 | 700 | 5.3 | 60196 | 7 | AAD55810 | Aad55810 Micromono |
| 27 | 689 | 5.2 | 427 | 5 | AAF67220 | Aaf67220 Novel hum |
| 28 | 689 | 5.2 | 110000 | 4 | AAI99682 | Aai99682 08 |
| 29 | 685 | 5.2 | 30690 | 4 | AAH79277 | Aah79277 Streptomy |
| 30 | 685 | 5.2 | 110000 | 4 | AAI99683 | Aai99683 08 |
| 31 | 683 | 5.2 | 30690 | 3 | AAA92301 | Aaa92301 S. avermi |
| 32 | 670 | 5.1 | 44377 | 2 | AAT80414 | Aat80414 Platenoli |
| 33 | 670 | 5.1 | 44377 | 2 | AAT80414 | Aat80414 Platenoli |
| 34 | 669 | 5.1 | 75236 | 7 | ABV75557 | Abv75557 Saccharop |
| 35 | 668.5 | 5.1 | 110000 | 4 | AAI99683 | Aai99683 27 |
| 36 | 666 | 5.0 | 58857 | 3 | AAAS8471 | Aaa8471 Nucleotid |
| 37 | 665 | 5.0 | 110000 | 4 | AAI99682 | Aai99682 27 |
| 38 | 664 | 5.0 | 110000 | 4 | AAI99682 | Aai99682 28 |
| 39 | 662.5 | 5.0 | 90600 | 6 | ABQ78872 | Abq78872 S. roseos |
| 40 | 661.5 | 5.0 | 43280 | 2 | AAT80413 | Aat80413 Ty lactone |
| 41 | 658.5 | 5.0 | 110000 | 4 | AAI99683 | Aai99683 39 |
| 42 | 654.5 | 5.0 | 113193 | 7 | AA054645 | Aad54645 Streptomy |
| 43 | 652 | 4.9 | 11358 | 9 | AD26983 | Adc26983 Sorangium |
| 44 | 652 | 4.9 | 13987 | 2 | AAT80415 | Aat80415 Hybrid sr |
| 45 | 651 | 4.9 | 110000 | 4 | AAI99683 | Aai99683 37 |
| 46 | 648.5 | 4.9 | 50937 | 3 | AAA09469 | Aaa09469 Streptoco |
| 47 | 647.5 | 4.9 | 65140 | 4 | AAI17184 | Aai17184 Streptomy |
| 48 | 647.5 | 4.9 | 125401 | 4 | AAI17186 | Aai17186 Streptomy |
| 49 | 646 | 4.9 | 110000 | 4 | AAI99682 | Aai99682 39 |
| 50 | 645 | 4.9 | 381 | 8 | ACH20441 | Ach20441 Human adu |
| 51 | 642 | 4.9 | 24379 | 2 | AAI93095 | Aai93095 Streptomy |
| 52 | 642 | 4.9 | 24379 | 2 | AAI93095 | Aai93095 Streptomy |
| 53 | 641 | 4.9 | 28598 | 2 | AAT06769 | Aat06769 Sorangium |
| 54 | 641 | 4.9 | 28598 | 2 | AAT89956 | Aat89956 Sorangium |
| 55 | 641 | 4.9 | 28598 | 2 | AAV75299 | Aav75299 DNA seque |
| 56 | 641 | 4.9 | 49377 | 3 | AAV05287 | Aav05287 The sorap |
| 57 | 639.5 | 4.8 | 31422 | 3 | AAA92302 | Aaa92302 S. avermi |
| 58 | 639.5 | 4.8 | 31422 | 4 | AAH79278 | Aah79278 Streptomy |
| 59 | 638 | 4.8 | 110000 | 4 | AAI99682 | Aai99682 37 |
| 60 | 636.5 | 4.8 | 110000 | 4 | AAI99683 | Aai99683 03 |
| 61 | 631.5 | 4.8 | 50000 | 4 | AAF88313 | Aaf88313 S. spinos |
| 62 | 631.5 | 4.8 | 50000 | 4 | AAF88316 | Aaf88316 S. spinos |
| 63 | 631.5 | 4.8 | 80161 | 2 | AAZ21501 | Aaz21501 DNA fragm |
| 64 | 626.5 | 4.7 | 110000 | 4 | AAI99682 | Aai99682 03 |
| 65 | 624.5 | 4.7 | 68750 | 3 | AAZ55887 | Aaz55887 Sorangium |
| 66 | 622.5 | 4.7 | 71989 | 3 | AAA29349 | Aaa29349 Sorangium |
| 67 | 622 | 4.7 | 11238 | 7 | AAD55817 | Aad55817 Micromono |
| 68 | 620.5 | 4.7 | 82746 | 7 | AAI61224 | Aai61224 Actinosyn |
| 69 | 619 | 4.7 | 47981 | 4 | AAF30757 | Aaf30757 Micromono |
| 70 | 618 | 4.7 | 110000 | 4 | AAI99682 | Aai99682 12 |
| 71 | 616.5 | 4.7 | 8438 | 2 | AAQ73500 | Aaq73500 DNA enco |
| 72 | 616 | 4.7 | 14775 | 2 | AAF88338 | Aaf88338 S. spinos |
| 73 | 615.5 | 4.7 | 15872 | 4 | AAT68715 | Aat68715 Streptomy |
| 74 | 614.5 | 4.7 | 103599 | 4 | ABX04971 | Abx04971 S. cinnam |
| 75 | 614.5 | 4.7 | 110000 | 4 | AAI99683 | Aai99683 06 |
| 76 | 613.5 | 4.6 | 110000 | 4 | AAI99683 | Aai99683 06 |
| 77 | 613.5 | 4.6 | 110000 | 4 | AAI99683 | Aai99683 12 |
| 78 | 612.5 | 4.6 | 82746 | 7 | AAI61224 | Aai61224 Actinosyn |
| 79 | 609.5 | 4.6 | 16767 | 7 | AAF88339 | Aaf88339 S. spinos |
| 80 | 608.5 | 4.6 | 16767 | 7 | AAF88339 | Aaf88339 S. spinos |
| 81 | 606 | 4.6 | 34094 | 2 | AAZ30163 | Aaz30163 Complete |